

A Multiscale, Hydrometeorological Forecast Evaluation of National Water Model Forecasts of the May 2018 Ellicott City, Maryland, Flood

FRANCESCA VITERBO,^{a,b} KELLY MAHONEY,^b LAURA READ,^c FERNANDO SALAS,^d
BRADFORD BATES,^e JASON ELLIOTT,^f BRIAN COSGROVE,^g AUBREY DUGGER,^c
DAVID GOCHIS,^c AND ROBERT CIFELLI^b

^a CIRES, University of Colorado Boulder, Boulder, Colorado

^b NOAA/Physical Sciences Division, Boulder, Colorado

^c National Center for Atmospheric Research, Boulder, Colorado

^d NOAA/National Water Center, Tuscaloosa, Alabama

^e Lynker Technologies, NOAA/National Water Center, Tuscaloosa, Alabama

^f National Weather Service Baltimore/Washington Forecast Office, Sterling, Virginia

^g Office of Hydrologic Development, NOAA/National Weather Service, Silver Spring, Maryland

(Manuscript received 31 May 2019, in final form 10 January 2020)

ABSTRACT

The NOAA National Water Model (NWM) became operational in August 2016, producing the first ever real-time, distributed, continuous set of hydrologic forecasts over the continental United States (CONUS). This project uses integrated hydrometeorological assessment methods to investigate the utility of the NWM to predict catastrophic flooding associated with an extreme rainfall event that occurred in Ellicott City, Maryland, on 27–28 May 2018. Short-range forecasts (0–18-h lead time) from the NWM version 1.2 are explored, focusing on the quantitative precipitation forecast (QPF) forcing from the High-Resolution Rapid Refresh (HRRR) model and the corresponding NWM streamflow forecast. A comprehensive assessment of multiscale hydrometeorological processes are considered using a combination of object-based, grid-based, and hydrologic point-based verification. Results highlight the benefits and risks of using a distributed hydrologic modeling tool such as the NWM to connect operational CONUS-scale atmospheric forcings to local impact predictions. For the Ellicott City event, reasonably skillful QPF in several HRRR model forecast cycles produced NWM streamflow forecasts in the small Ellicott City basin that were suggestive of flash flood potential. In larger surrounding basins, the NWM streamflow response was more complex, and errors were found to be governed by both hydrologic process representation, as well as forcing errors. The integrated, hydrometeorological multi-scale analysis method demonstrated here guides both research and ongoing model development efforts, along with providing user education and engagement to ultimately engender improved flash flood prediction.

1. Introduction

On 27 May 2018 Historic Ellicott City, Maryland, experienced for the second time in 22 months a “1-in-1000 chance” torrential rainfall event that brought catastrophic flood damage (nearly \$10.5 million, according to the NOAA StormData database) and one fatality. Heavy rainfall, totaling between 150 and 300 mm in the heaviest 6-h band, fell in central Maryland (Fig. 1). The flood damaged many buildings and vehicles. Some roads were washed out, and land erosion and localized

landslides were reported in the region (National Weather Service 2018).

Forecasting such rare, intense, small-scale flood events remains a salient challenge in weather and hydrologic prediction. This event is thus examined as a case study that exemplifies many of the challenges typically faced in forecasting extreme flash floods. While there are many existing models and forecasting practices established for flash flood prediction, here we specifically examine the performance and future potential of forecasting events such as the Ellicott City 2018 flood using the operational National Oceanic and Atmospheric Administration (NOAA) National Water Model (NWM).

Flash flood forecasting has benefitted from a number of observational and modeling advancements over

Corresponding author: Francesca Viterbo, francesca.viterbo@noaa.gov

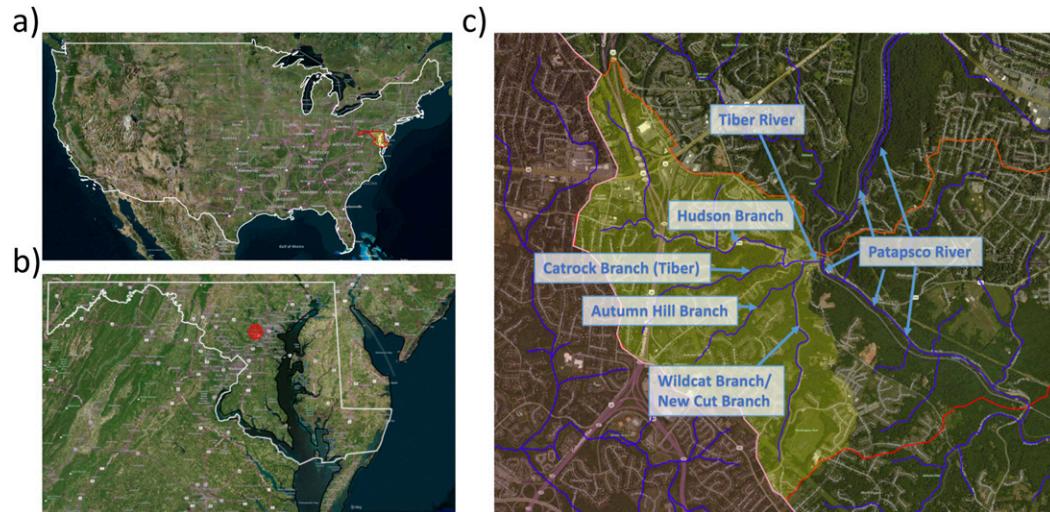


FIG. 1. Study area and Ellicott City watershed. (a) The position of Maryland in the continental United States, (b) the position of Ellicott City in the state of Maryland, and (c) the 9.5-km² watershed that encloses the Ellicott City downtown area.

recent decades, but a progressive focus toward smaller spatial scales presents new challenges at the interface of hydrology and meteorology. Streamflow forecasts strongly depend on the quality of the meteorological inputs that are provided to the hydrological model; meteorological forcing, model structure, and parametric uncertainty all affect the final streamflow forecast. Considering the vast suite of uncertainties in both the atmospheric forcings and those intrinsic to the hydrological modeling framework, understanding error propagation through the forecast chain is challenging, but essential to improving predictions (Collier 2007).

The NWM officially became an operational hydrologic forecasting system in August 2016, producing real-time, high-resolution, distributed hydrologic forecasts for the continental United States (CONUS), its contributing basins, and Hawaii. The initial objective of the NWM is to complement hydrologic guidance at current National Weather Service (NWS) river forecast locations and significantly expand forecast coverage and type in underserved locations. At the time of writing, few studies have been published regarding the usage and performance of this new modeling framework, especially in conjunction with a detailed study on the meteorological forecast performance. Given the complexity and scale of this hydrologic forecast system, ongoing evaluation is crucial to guide future improvements.

In this work the 2018 Ellicott City flash flood event is studied with a particular focus on short-range (0–18 h) streamflow forecasts from the NWM, forced by the quantitative precipitation forecasts (QPFs) from the High-Resolution Rapid Refresh (HRRR)

model (Benjamin et al. 2016). The short-range forecast configuration is intended to focus on small spatial and temporal scales, which still represent one of the most challenging forecast targets for operational numerical weather prediction (Collier 2007). It is thus also the scale that would benefit greatly from the NWM's high resolution (2.7 million river segments for channel routing; 1-km land surface and 250-m surface routing grids) and spatially distributed forecast capabilities in potential flash flood conditions. Though the NWM is implemented at a higher resolution than most other current CONUS operational forecast models, the Ellicott City watershed represents a particular challenge in testing the skill at such a small scale: the Ellicott City watershed (9.5 km²) barely covers 9 grid elements over the total 17 000 000 1-km NWM land grid elements. Coupled with the fact that the CONUS-scale, distributed NWM differs from current operational hydrologic models that are run offline and have been locally calibrated over long periods, a successful NWM forecast of an event on this scale would seem almost serendipitous if it were able to “nail” a highly localized flood forecast of this magnitude.

The objective of this paper is to investigate the short term predictive capability of NWM in a flash flood prone area, while attempting to understand the relative uncertainties attributable to precipitation forcings versus hydrologic process representation. While a complete quantitative separation of model uncertainty within the flood forecasting chain is beyond the scope of this single case study, we demonstrate a method that can guide future studies of similar events. One anticipated outcome of this work is to use the findings to inform

ongoing NWM model assessment and future version improvement, ultimately resulting in improved local forecast guidance.

The study addresses these targets using a multiscale approach and a combination of verification methods to combine a meteorological mesoscale QPF evaluation, an analysis of the small catchment hydrology response, and an urban-scale flood inundation analysis. The study region is described in [section 2](#), and specific event details are provided in [section 3](#). Data and methods are described in [section 4](#), while the results are discussed in [section 5](#). Final conclusions are drawn in [section 6](#).

2. Study region

Ellicott City, Maryland, was founded in 1772 as a mill town on the Patapsco River. The town has flooded multiple times from the larger Patapsco River that runs alongside the historic downtown district, as well as from the Tiber River, which runs through it ([Fig. 1](#)). The Tiber watershed is quite small, on the order of several square kilometers, but consists of four small branches: Tiber Branch, New Cut Branch, Autumn Hill Branch, and Hudson Branch, which converge near Main Street in Ellicott City to form the Tiber ([Fig. 1c](#)). This watershed contains rugged terrain with sharp gradients, sometimes dropping more than 30-m elevation over 100 m of horizontal distance ($\sim 17^\circ$ slope).

The Patapsco River Lower North Branch watershed (where Ellicott City is located) is about 306 km² and contains a variety of land uses. Land cover is 49% developed, 41% forest area, and 10% agricultural. The general tendency in the watershed is to have developed lands near the headwaters and upstream locations, while forested land is in downstream areas ([Maryland Department of Natural Resources Watershed Services 2005](#)). Impervious surfaces (including roads and rooftops) are present in a high percentage in most of the watershed, further preventing stormwater from infiltrating the ground. The Tiber–Hudson watershed, which contains Ellicott City, has the highest percentage of impervious surfaces (about 25%–30%), and this significantly impacts local stream drainage.

Most of the soils in the area are characterized as stony or shallow bedrock. In the Tiber Branch there is a high concentration of hydric soils with limited infiltration capacity ([Maryland Department of Natural Resources Watershed Services 2005](#)). The remaining portion of the Ellicott City basin is characterized by soils with generally good infiltration capacity; however, the topographic slope often exceeds 8%, rendering runoff too fast to infiltrate, despite permeable soil types. In summary, the area is a mix of hydric soils (highly saturated), stony soils

(limited infiltration capabilities), and complex topography; these factors enhance surface runoff in this region and create a flashy rainfall–runoff response.

Because of the combination of these regional soil characteristics and Ellicott City's location in a valley at the confluence of two major creeks, the area has been affected by a number of major floods. While many historic floods have originated from the Patapsco River overflowing its banks and flooding upstream areas (e.g., Hurricane Agnes in 1972 along with many other events), the 2016 and 2018 floods were instead caused by rapid upstream confluence of water caused by intense rainfall over the steep upstream branches of the Tiber watershed ([Fig. 1](#)). These last two flood events happened within a 22-month period (30 July 2016 and 27 May 2018), each causing catastrophic flood damage. Though many similarities exist with the 2016 event, this study will focus on the May 2018 event (as the NWM was not yet operational in July 2016.)

3. Event overview

The synoptic weather pattern on 27 May 2018 was dominated by high pressure over much of the mid-Atlantic U.S. region, allowing high temperatures and record-breaking precipitable water values (nearly 50 mm; [Fig. 2a](#)) to produce convective available potential energy (CAPE) values between 1000 and 3000 J kg⁻¹. A line of thunderstorms initiated along a surge of relatively colder offshore air from the North Atlantic Ocean moving from northeast to southwest ([Fig. 2b](#)); the so-called “backdoor cold front” was marked by low pressure at the surface that further enhanced convergence of moist, unstable air as the front moved across Maryland. As the line of thunderstorms formed and moved east-southeastward, midlevel winds remained westerly, creating training cells over central Maryland from approximately 1800 UTC 27 May to 0000 UTC 28 May 2018 ([Fig. 3](#)).

The training behavior of the convective cells led to persistent heavy rainfall over the region from 1800 to 2300 UTC 27 May, punctuated by intensified pulses as multiple low-level boundaries triggered and maintained precipitation for several hours. Over Ellicott City itself, precipitation peaked first between 1900 and 2000 UTC, after which there was a short lull before rainfall reached maximum intensity between 2120 and 2150 UTC. Surrounding cities in central Maryland also received significant rainfall, but the maximum event total accumulation was over the Ellicott City region, and according to station- and radar-based observations (discussed below), exceeded 200 mm ([Fig. 2c](#); [National Weather Service 2018](#)).

NWS forecasts recognized the potential for heavy rainfall several days in advance, with a Hazardous

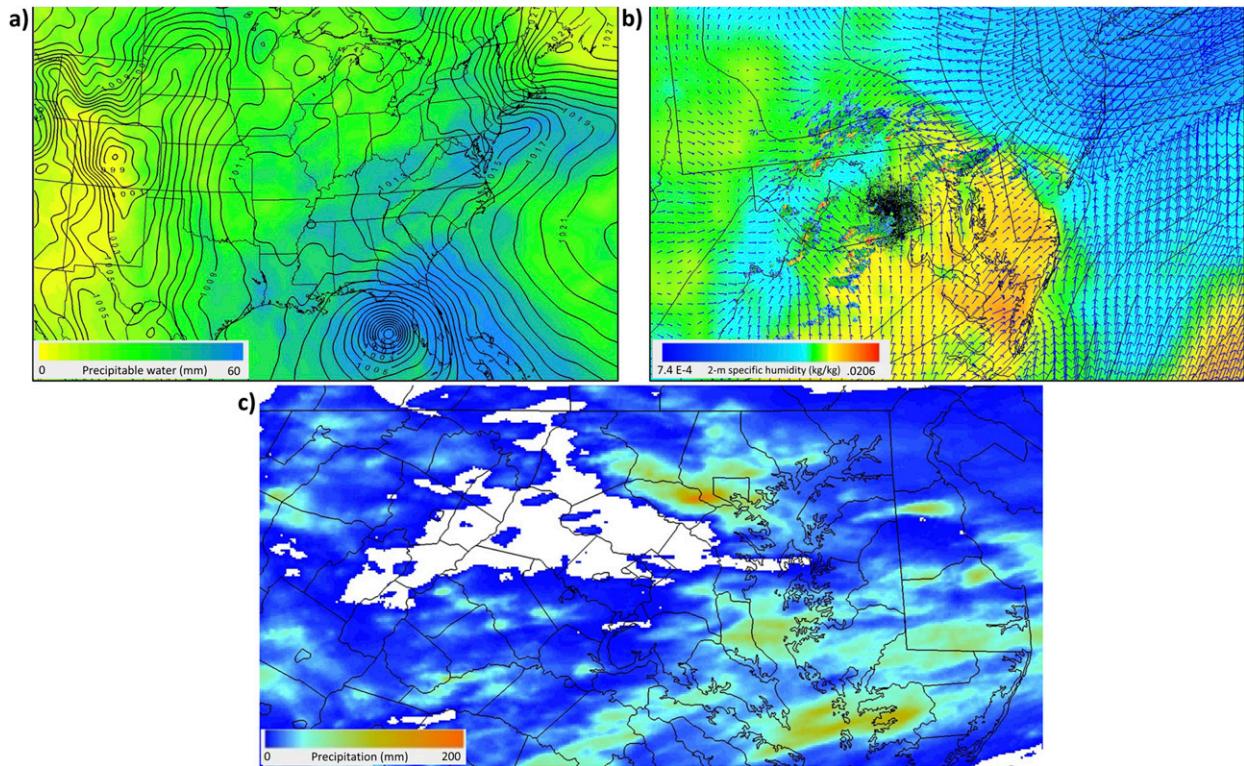


FIG. 2. Large-scale synoptic pattern and resulting rainfall from the backdoor cold front that affected Ellicott City on 27 May 2018. (a) Precipitable water (mm; as shaded in colorbar) and mean sea level pressure (black contours; interval 1 hPa, as labeled) from the NOAA Rapid Refresh (RAP) model analysis. (b) The 2-m specific humidity (kg kg^{-1} ; as shaded in colorbar), mean sea level pressure (black contours; interval 0.5 hPa), 10-m wind vectors from the RAP model analysis valid 1800 UTC 27 May 2018; KLRX radar reflectivity overlaid valid 1804 UTC 27 May 2018. (c) MRMS 24-h observed precipitation analysis (mm; as shaded in colorbar) valid 1200 UTC 28 May 2018.

Weather Outlook issued 25 May citing “thunderstorms capable of producing locally heavy rain may lead to isolated incidents of flooding on Saturday [26 May] and Sunday [27 May].” On the morning of 27 May a flash flood watch was issued for the region, progressing to a flood warning issued at 1919 UTC, then a flash flood warning at 2026 UTC, and finally culminating in a flash flood emergency declaration at 2040 UTC.

Rapid river rises and swift increases in river stage were observed not only in Ellicott City, but also in nearby locations. The Patapsco River at Ellicott City peaked around 2230 UTC, reaching a level close to the moderate flood stage. Due to heavy rain upstream and in surrounding areas, high water volume was already present in the Patapsco; this prevented Ellicott City’s flood waters (from both the Tiber River and from the urban land surface) from draining effectively into the Patapsco (Fig. 1). Other streamflow gauges along the Patapsco (near Catonsville and Elkridge) approached or exceeded moderate flood level thresholds. Flooding was also widespread along the Patuxent and Gwynns

Falls catchments in Catonsville, Baltimore City, and the Dundalk and Perry Hall areas (National Weather Service 2018). While the Ellicott City basin is ungauged, indirect, postmortem peak flow estimates by the USGS calculated peak discharge on the Hudson, Catrock (Tiber), and Autumn Hill/New Cut branches to be 84.95, 76.45, and $174.43 \text{ m}^3 \text{ s}^{-1}$ respectively, all exceeding the estimates from the 2016 flooding event.

4. Data and methods

The 2018 Ellicott City flood is examined using a multiscale, multiprocess approach, starting from the meteorological forcing inputs of the NWM at 1 km, to the runoff and streamflow response at the hydrologic catchment scale. The workflow described in Fig. 4 includes a detailed QPF analysis of the short-range NWM forcings from the HRRR model, NWM streamflow forecast verification for the Ellicott City area and surrounding catchments, and flood inundation assessment at the urban/local scale in historic downtown Ellicott City where the greatest flood impacts occurred.

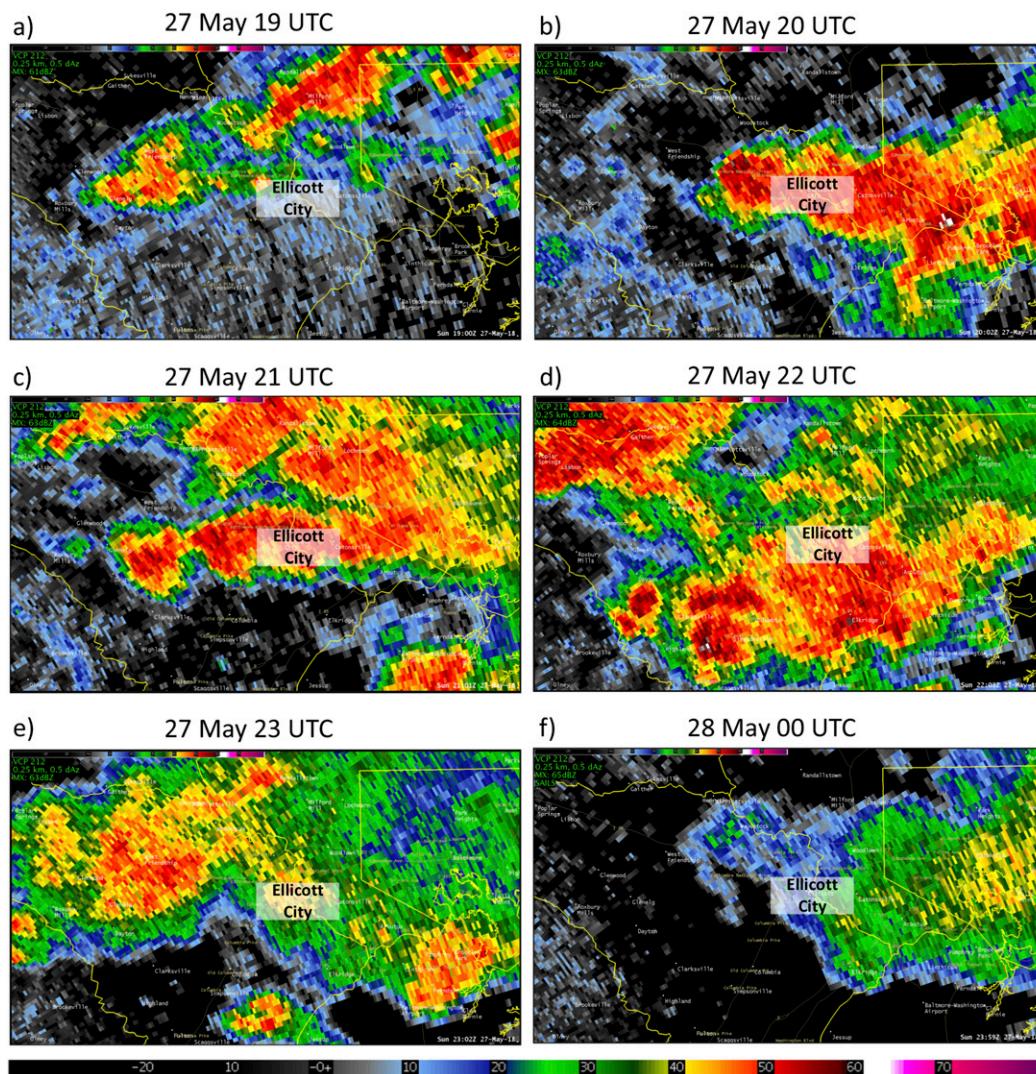


FIG. 3. Hourly evolution of radar reflectivity from the KLWX NEXRAD at (or the closest instant to) (a) 1900, (b) 2000, (c) 2100, (d) 2200, and (e) 2300 UTC 27 May and (f) 0000 UTC 28 May. The radar reflectivity values corresponding to the color shading are shown in the legend at the bottom of the figure.

A detailed description of the study methods is provided in the following paragraphs.

a. QPF verification

The QPF analysis compares the short-term precipitation forecasts produced by the NCEP HRRR model with the quantitative precipitation estimation (QPE) product from the gauge-corrected Multi-Radar Multi-Sensor (MRMS) analysis (Zhang et al. 2011, 2014). The high temporal (2 min) and spatial (1 km) resolution in addition to ample surface observations and complete radar coverage in this region render the MRMS product well suited for flash flood forecasting applications, particularly when rainfall–runoff response times are subhourly (e.g., in steep headwater basins and in urban

areas like Ellicott City) (Gourley et al. 2017). Verification analyses and statistics are produced using the Model Evaluation Tools (MET) developed by the NCAR Developmental Testbed Center (Developmental Testbed Center 2018).

The HRRR model (Benjamin et al. 2016) is the meteorological model used to force the NWM short-range cycle, and at the time of the 2018 Ellicott City flood, produced hourly forecasts from 0- to 18-h lead time. The HRRR model uses 3-km grid spacing and is built upon the Advanced Research core of the Weather Research and Forecasting Model (Skamarock et al. 2008). It is an hourly updated, cloud-resolving model for operational use that runs over the CONUS domain. HRRR model version 2 was operational at the time of the 2018 Ellicott

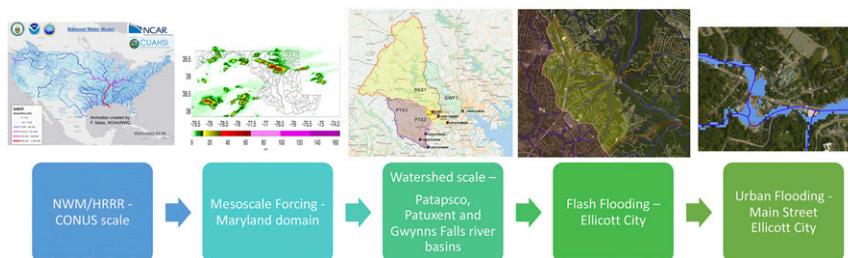


FIG. 4. Multiscale methodology from larger (CONUS) to smaller scales (urban flooding).

City flood and is thus the version examined here. The study analyzes the HRRR forecasts cycles from 1300 to 1900 UTC on 27 May, which captures the hours leading up to and including the most intense precipitation (2000 UTC 27 May–0000 UTC 28 May).

The QPF analysis is performed over a region including Maryland and portions of surrounding states to capture the main mesoscale characteristics of the event (Fig. 1b). QPF is also evaluated over key watersheds in order to consider the precipitation forecast (and all of its possible displacement, intensity, and timing errors) through the more hydrologically relevant concept of mean areal precipitation (which connects more directly to NWM outputs and hydrologic impacts). Figure 5b shows the watershed locations, and Table 1 provides a summary of the main characteristics of the selected sites where the streamflow evaluation is performed. Peak discharge per unit area (defined as the peak flow over the event divided by the contributing watershed area) is also provided for each site as a reference point using USGS observed peak flows where available.

Hourly HRRR precipitation is verified against the MRMS QPE by first regridding the HRRR data at its native resolution (3-km grid spacing) to the same MRMS grid (1-km grid spacing) and then bilinearly interpolating the data with a 2×2 interpolation box. The forecast cycles are then verified using a suite of MET gridded and object-based verification tools. Standard verification statistics [e.g., bias (mean average error), RMSE, correlation] are computed hourly over matched QPE–QPF grid points, as well as for accumulated statistics over the full 18-h forecast cycle and summed gridded fields.

Finally, mesoscale storm characteristics are evaluated with an object-based verification method using the Method for Object-Based Diagnostic Evaluation (MODE) tool from the MET package. Object-based verification is important for high-resolution forecast evaluation as it helps to avoid the “double penalty” problem whereby an overall skillful, but slightly spatially offset, model forecast is penalized for under- (over) prediction where precipitation was (not) observed (Rossa et al. 2008).

Here we aim to also recognize where forecasts with generally accurate intensity, timing, and only slightly displaced location characteristics still may offer value to a forecaster. To this end, MODE identifies objects that a human eye would recognize as a “region of interest” in both model and observation fields using two main parameters: the convolution radius and the convolution threshold. The convolution radius specifies the radius of the convolution filter that is applied to the precipitation field (expressed in number of grid boxes), while the convolution threshold applies a filter on the intensity of the fields that are identified within the objects. The convolution radius and the convolution threshold are set by the user through several iterative experiments. In this study the convolution radius and the convolution threshold are set to 8 grid boxes and 10 mm over 1-h rainfall accumulation, respectively. MODE calculates attributes of interest for each of the identified objects (e.g., area, intensity, shape of the

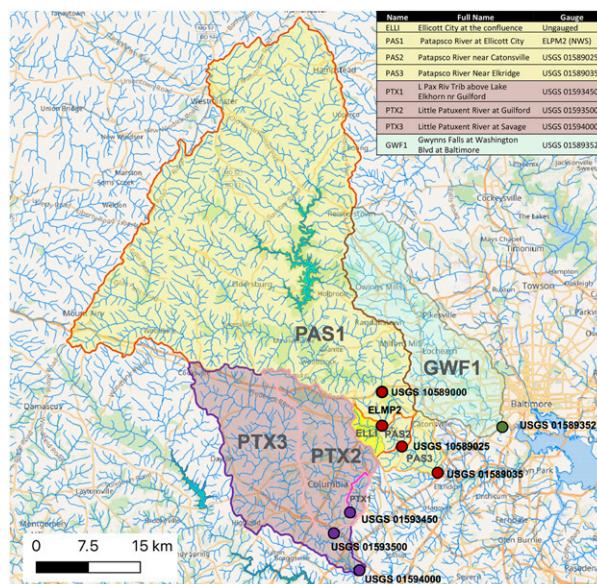


FIG. 5. Watersheds (polygon boundaries) examined in the study area with USGS gauge locations (points) and the NWM flow line network (blue lines).

TABLE 1. Main characteristics of the watersheds in the study and peak discharge per unit area information regarding the watersheds in exams for the event. The ELLI watershed peak discharge per unit area has an asterisk because it was calculated using the NWM analysis values, since it is an ungauged site.

Name	Full name	Gauge	River	Drainage area (km ²)	Peak discharge per unit area (m ³ s km ⁻²)
ELLI	Ellicott City at the confluence	Ungauged	Tiber	9.5	13.09*
PAS1	Patapsco River at Ellicott City	ELPM2 (NWS)	Patapsco	761.5	0.70
PAS2	Patapsco River near Catonsville	USGS 01589025	Patapsco	780	0.83
PAS3	Patapsco River near Elkridge	USGS 01589035	Patapsco	808.1	0.93
PTX1	Little Patuxent River tributary above Lake Elkhorn near Guilford	USGS 01593450	Little Patuxent	6.4	4.04
PTX2	Little Patuxent River at Guilford	USGS 01593500	Little Patuxent	98.4	2.17
PTX3	Little Patuxent River at Savage	USGS 01594000	Little Patuxent	254.9	1.24
GWF1	Gwynns Falls at Washington Blvd at Baltimore	USGS 01589352	Gwynn Falls	170.7	2.62

identified object, position of the centroid) and merges and matches accordingly objects that have similar characteristics within the same field (defined as “merging”) and across the model and the observation field (defined as “matching”). Finally, each of the corresponding matching and merging objects are “paired” and used by MODE to calculate the statistics.

Because the attributes of complexity and intensity are analyzed in the results discussion in section 5a(2), a brief definition of these attributes is given hereinafter. Complexity is defined as the area of the object to the area of its convex hull (defined as the smallest convex set that contains a set of points or an object), while the intensity is considered as the intensity of the field inside the object and it can be a sum of all the intensities in the object (intensity sum) or the percentile of the intensity.

Since MET offers many possible evaluation metrics, we select a subset most relevant for the QPF analysis and refer the reader to the MET literature (Davis et al. 2006; Brown et al. 2009; Bullock et al. 2016) and to the MET user guide (Developmental Testbed Center 2018) for additional information.

b. NWM evaluation

NWM streamflow forecasts are compared with available USGS and NWS streamflow observations to understand the runoff response from the HRRR meteorological forcing for this event. We note that while comparing NWM forecast guidance to existing operational hydrologic forecast guidance would be desirable, the catchments that feed into Ellicott City itself are ungauged and do not include any official National Weather Service River Forecast Center forecast points. Therefore, the NWM is unique in its ability to be a potential forecast tool for both streamflow and gridded land surface variables in this area. This presents an opportunity and a challenge, as relative to local hydrologic models with rich observational histories from which to

locally calibrate and benchmark forecasts, there is a significant challenge in communicating how one uses NWM forecasts as guidance in locations where there simply was not any prior.

1) THE NATIONAL WATER MODEL (NWM)

The NWM modeling framework is based on a configuration of the Weather Research and Forecasting (WRF)-Hydro modeling system (Gochis et al. 2015), which combines the Noah-MP land surface model (Niu et al. 2011) at 1 km to compute infiltration and exfiltration, with a diffusive wave overland routing formulation (Julien et al. 1995; Ogden 1998) across the land at 250-m resolution; the channel routing module then receives water from adjacent catchments and uses the Muskingum–Cunge method (Garbrecht and Brunner 1991) to route water down a flow network based on the National Hydrography Dataset version 2 (NHDPlusv2) (McKay et al. 2012). The NWM version 1.2 analyzed in this study uses the National Land Cover Database 2011 (NLCD 2011; Homer et al. 2015), upscaled from 30 m to a 1-km grid by taking the major class within that pixel, and takes the soil characteristics from State Soil Geographic (STATSGO) 2006 (Wolock 1997), downscaled with the same method from a resolution of 1° × 2° to 1 km.

As of version 1.2, the NWM simulates streamflow and other hydrologic variables in four different temporal cycles: analysis and assimilation cycle (AnA), and short- (0–18 h), medium- (0–10 days), and long-range (0–30 day, ensemble) forecast configurations (Office of Water Prediction 2016). For the Ellicott City case study, we focus on the short-range (0–18 h) forecast, and also use the AnA cycle, which includes streamflow observations via assimilation and meteorological forcings from the gauge-corrected MRMS, to provide a sense of streamflow in ungauged areas. The AnA cycle is run hourly to produce a real-time analysis of current

hydrological conditions at the surface and near-surface over the CONUS domain. The AnA cycle also provides the initialization states for the short, medium and long-range forecasts. Streamflow observations are assimilated hourly at about 7000 USGS gauge locations depending on the availability and quality of the observed data. About 1260 lakes and reservoirs are included in the model version analyzed here, using a simplified level-pool scheme parameterization. The hourly short-range forecasts evaluated in this work were forced by the HRRR model and initialized using the AnA configuration. For further reading on the NWM, the reader is referred to [Office of Water Prediction \(2016\)](#) and references therein.

2) STREAMFLOW EVALUATION

Streamflow observations are provided by USGS stream gauges and available NWS local streamflow locations. [Figure 5](#) illustrates the location of the gauge locations most impacted by the 2018 Ellicott City flood, as well as relevant upstream watersheds, and [Table 1](#) lists the main watershed characteristics and illustrates the peak discharge per unit area observed during the event. These watersheds are also used for the catchment-scale QPF analysis.

The short-range NWM streamflow forecasts from the hourly model start times from 1300 to 1900 UTC 27 May 2018 are compared to gauge observations.

Forecast skill is assessed both by qualitative hydrograph comparison, and quantitative forecast skill score evaluation [e.g., correlation ([Hazewinkel 2013](#)), Nash–Sutcliffe efficiency (NSE; [Nash and Sutcliffe 1970](#)), bias ([Van der Vaart 1961](#)), peak flow ([Mockus 1964](#)), time of the peak ([Mockus 1964](#)), and streamflow volume ([Mockus 1964](#))]. Runoff results are compared to the mean average precipitation evaluation. This comparison evaluates the combined (meteorological and hydrological) forecast skill of each cycle and illuminates times and locations propagation of possible QPF errors into the streamflow response.

c. Flood inundation analysis

Flood inundation maps in the urban Ellicott City area were derived from the NWM using the Height Above Nearest Drainage (HAND) method ([Rodda 2005](#); [Rennó et al. 2008](#); [Nobre et al. 2016](#); [Tesfa et al. 2011](#); [Zheng et al. 2018](#)). The methodology is very similar to [Liu et al. \(2018\)](#) that leveraged a 10-m HAND grid derived from the USGS 10-m digital elevation model. In general, as described in [Nobre et al. \(2016\)](#), the HAND method identifies the difference in elevation between each land surface grid point and the stream bed cell to which it drains. This normalized relative elevation

grid (i.e., the HAND grid) allows a flood stage value along a given river segment to be mapped to the catchment, essentially “flooding” all of the cells that are at or below the flood stage value (i.e., the HAND value), resulting in inundation extent along the channel.

The experimental NWM–HAND method used in this study uses operational NWM streamflow output (i.e., water already in the channel), converted to stage, to simulate inundation on the land surface. NWM streamflow is translated to stage, using synthetic rating curves along each individual stream segment. Channel properties such as volume, cross-sectional area, wetted perimeter, and hydraulic radius are derived using this method, assuming a Manning’s value and using NHDPlus channel slope. While some uncertainty is introduced by these assumptions, only inundation extent is considered here (versus water depth modeling), and so we use it as an initial demonstration of capability. The reader is referred to [Zheng et al. \(2018\)](#) for the method of derivation of synthetic rating curves.

The HAND method is applied to the short-range NWM forecast cycles from 1300 to 1900 UTC 27 May. The maximum inundation extent for a single forecast (e.g., 1300 UTC) is compared to the maximum inundation extent derived from the AnA files over the same period, herein assuming that the NWM analysis can be considered as a sufficient proxy to represent the true conditions. To support this assumption, the flood maps are compared to 12 street cameras that were placed in historical downtown Ellicott City by an association of citizens after the 2016 flood (Ellicott City Unilux Camera Network; [Peters 2019](#)). Camera locations are shown in [Fig. 6](#). The cameras also allow for evaluation of the temporal evolution of the flooding inundation extent.

5. Results and discussion

a. Mesoscale QPF verification

1) GRIDDED VERIFICATION

In this section precipitation is evaluated by comparing the HRRR forecast cycles from 1300 to 1900 UTC 27 May with MRMS QPE. [Figure 7](#) shows mean precipitation error over the Maryland region calculated over the full forecast cycle length (18 h) using the MET series analysis tool ([section 4a](#)). Across all forecast cycles, the accumulated absolute bias shows an underestimation of precipitation over Ellicott City. The spatial pattern also highlights areas of overestimation; side-by-side couplets of over and underforecast of the QPF suggest displacement errors more than a systematic QPF



FIG. 6. Position of the street cameras used in the flood inundation map comparison.

underestimation over the area. While there is clearly run-to-run variability as well as some smoothing of hourly displacement errors from this statewide run-total perspective, relative to the large amounts of precipitation (greater than 200 mm) that led to the flood event, the HRRR forecasts generally appear to be skillful at the mesoscale, depicting considerable likelihood of an intense precipitation event in the central Maryland area.

Zooming in to a watershed scale centered around Ellicott City, and focusing on the six hours (from 1800 UTC 27 May to 0000 UTC 28 May) encompassing the flood event, the MRMS observations (Fig. 8a) indicates the most intense 6-h rainfall over the Ellicott City watershed (ELLI), the lower part of the Patapsco (PAS1, PAS2, and PAS3), lower GWF1, and upper Patuxent watersheds. The 1300 (Fig. 8b), 1500 (Fig. 8d), and 1600 UTC (Fig. 8e) HRRR cycles reveal eastward and northward spatial displacements of the heaviest precipitation totals, while the 1700 UTC cycle (Fig. 8f) displaces 6-h rainfall maxima both north and south. The 1800 UTC cycle (Fig. 8g) exhibits the closest representation to MRMS and the most similar spatial orientation, even if still underestimating QPF over the Ellicott City watershed (ELLI). The 1400 UTC cycle (Fig. 8c) produces the salient features of the heavy precipitation, even if the pattern of heavy precipitation extends too far into the upper reaches of PAS1. Finally, the 1900 UTC cycle (Fig. 8h) has similar characteristics of the 1800 UTC cycle in terms of spatial orientation but overestimates precipitation on the upper Patapsco and places the most intense rainfall core too far south.

In Figs. 9 and 6h (1800 UTC 27 May–0000 UTC 28 May) HRRR QPF is aggregated at the single watershed scale, and the mean areal precipitation across each basin is compared the same quantities from MRMS. This analysis confirms HRRR's underestimation over ELLI

across all HRRR cycles, with the lowest bias seen in the 1800 UTC cycle. Across the other watersheds, the mean areal precipitation biases are variable in sign, and vary with HRRR forecast cycle, but overall the 1800 UTC cycle performs best, particularly in the areas of most intense rainfall.

2) OBJECT-BASED VERIFICATION

Particularly for hydrologic impact applications, small spatial displacements in precipitation can matter a great deal, and such displacements can be difficult to discern from standard grid-based verification methods. For this reason, we also evaluate storm and precipitation evolution in space and time using MET's MODE object-based verification tool. An example of possible displacement that has significant effects for hydrological application is given by the MODE analysis in Fig. 10, comparing the 1300 UTC cycle with MRMS at 2200 UTC (Figs. 10a,b): the HRRR forecasts intense and localized precipitation with similar mesoscale characteristics to those observed, but due to displacement error, precipitation does not fall in any of the watersheds that received flooding. MODE pairs an observed precipitation "object" with the most appropriate precipitation forecast "object," as shown in Figs. 10a1 and 10b2. The same color (or object ID, as defined in the MODE tool) within the forecast or observation field indicates that the two objects, even if spatially separate, are merged in the same merged object, due to similar characteristics. When the same color (or object ID) is matching across the forecasts and observations fields, it means that MODE has paired the forecasted object (or merged objects) with the corresponding(s) observed object (merged objects).

Figure 11 summarizes the temporal evolution of the matched observed and forecasted objects valid from 2000 UTC 27 May to 0000 UTC 28 May at hourly

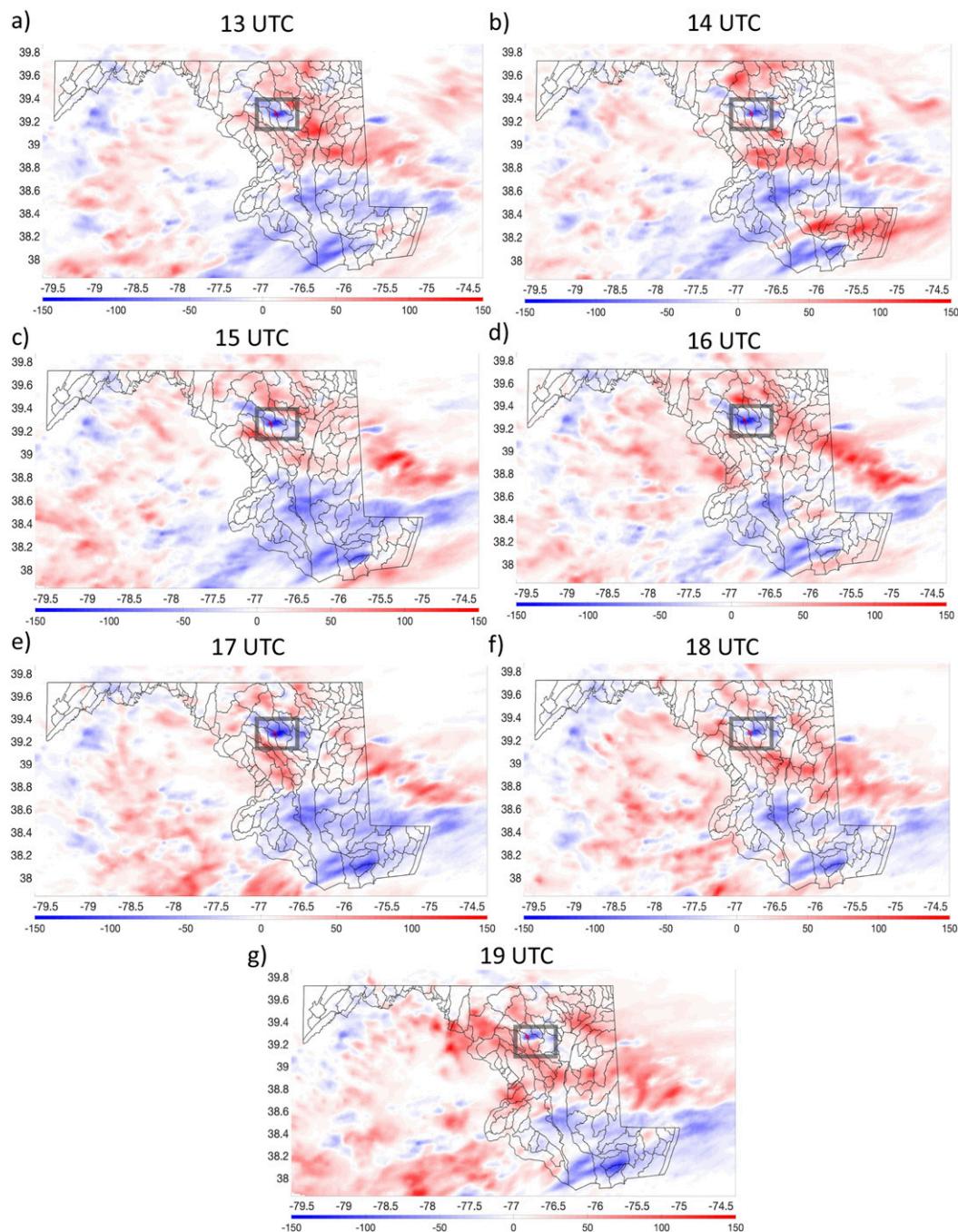


FIG. 7. (a)–(g) Accumulated absolute bias between HRRR cycles initialized from 1300 UTC to 1900 UTC 27 May 2018 and MRMS QPE, as obtained from the full 18-h duration of the short-range cycle using the Series Analysis tool from MET. The blue color means underestimation of the HRRR cycle, as compared to the MRMS, while the red color means overestimation of the HRRR cycle over the MRMS QPE. The gray rectangle highlights the HRRR underestimation compared to MRMS, over the area of the Ellicott City watershed (in bright red contours).

time steps from all the cycles from 1300 (Fig. 11a) to 1900 UTC (Fig. 11g). While grid-based aggregate analyses indicate systematic QPF underestimation over Ellicott City, if one takes an object-based approach and

instead follows the most intense precipitation objects in space and time, there is decidedly less error indicated, as forecast and observed values of total object intensity (Fig. 11, third column) and of the 90th percentile object

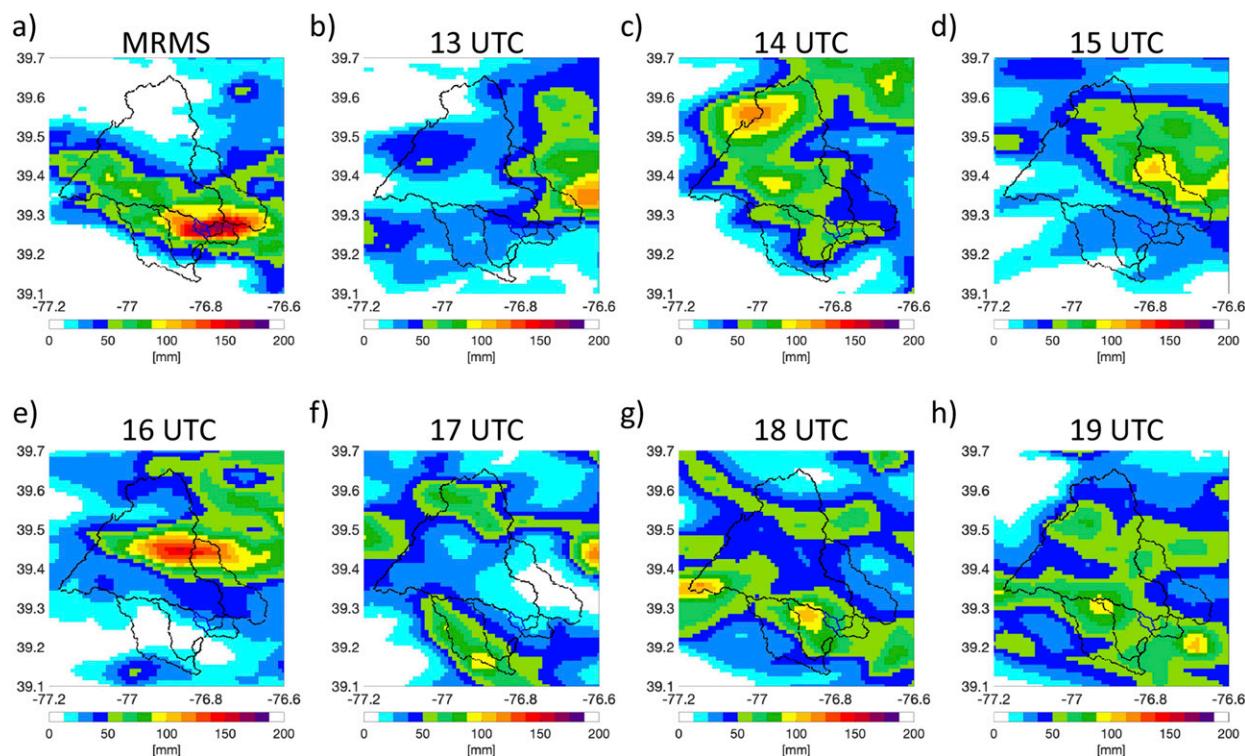


FIG. 8. The 6-h rainfall accumulation from 1800 to 0000 UTC for (a) MRMS OPE, (b) HRRR 1300 UTC, (c) HRRR 1400 UTC, (d) HRRR 1500 UTC, (e) HRRR 1600 UTC, (f) HRRR 1700 UTC, (g) HRRR 1800 UTC, and (h) HRRR 1900 UTC (note 5-h accumulation valid for available forecast hours from 1900 UTC to 0000 UTC). Watershed boundaries are indicated by black solid lines and correspond to watersheds names identified in Fig. 5.

intensity (Fig. 11, fourth column) are actually quite closely matched.

The 1800 UTC HRRR cycle performs well in terms of intensity and area over the full length of the forecast cycle, and the HRRR “complexity” (a measure of shape characteristics) appears quite skillful across most cycles. The 1300 and 1400 UTC HRRR cycles (Figs. 11a,b) also exhibit greater forecast–observed agreement in terms of area, complexity, and intensity. The 1500, 1600, 1700, and 1900 UTC (Figs. 11c, 11d, 11e, and 11g, respectively) cycles show an overestimation of size (area) (Fig. 11, first column) and summed intensity of the forecast objects (Fig. 11, third column), although focusing on the 90th percentile precipitation intensity demonstrates greater model skill (Fig. 11, fourth column). In summary, the MODE analysis suggests that HRRR is actually slightly overestimating the intensity of the main precipitation features (albeit spatially displaced) in most of the cycles, with relatively better performance in the 1800 and 1900 UTC cycles.

b. Watershed-scale evaluation

In this section, we combine the precipitation-only evaluation performed in section 5a with the NWM

short-range streamflow forecast. While Ellicott City bore the brunt of the event in terms of flood damage and loss of human life, surrounding watersheds also experienced significant flooding, as discussed in section 3. To facilitate the NWM assessment over the collection of watersheds described in Fig. 5 and Table 1, the results will be discussed in terms of moving down the network from small headwaters to larger drainage basins (and those within). In the small upstream watershed discussion [section 5b(1)], the results over Ellicott City (ELLI) and over the Little Patuxent River tributary above Lake Elkhorn near Guilford (PTX1) are discussed and summarized in Fig. 12. Because the ELLI watershed does not have streamflow observations and we can only compare the NWM to the analysis, a nearby small watershed with similar characteristics and size, PTX1, is used for comparison.

The three closing sections of the watersheds with relatively larger drainage areas were evaluated with respect to streamflow response [section 5b(2)]: the Gwynns Falls River at Washington Blvd. at Baltimore (GWF1), the Little Patuxent River at Savage (PTX3), and the Patapsco River near Elkridge (PAS3). Note that

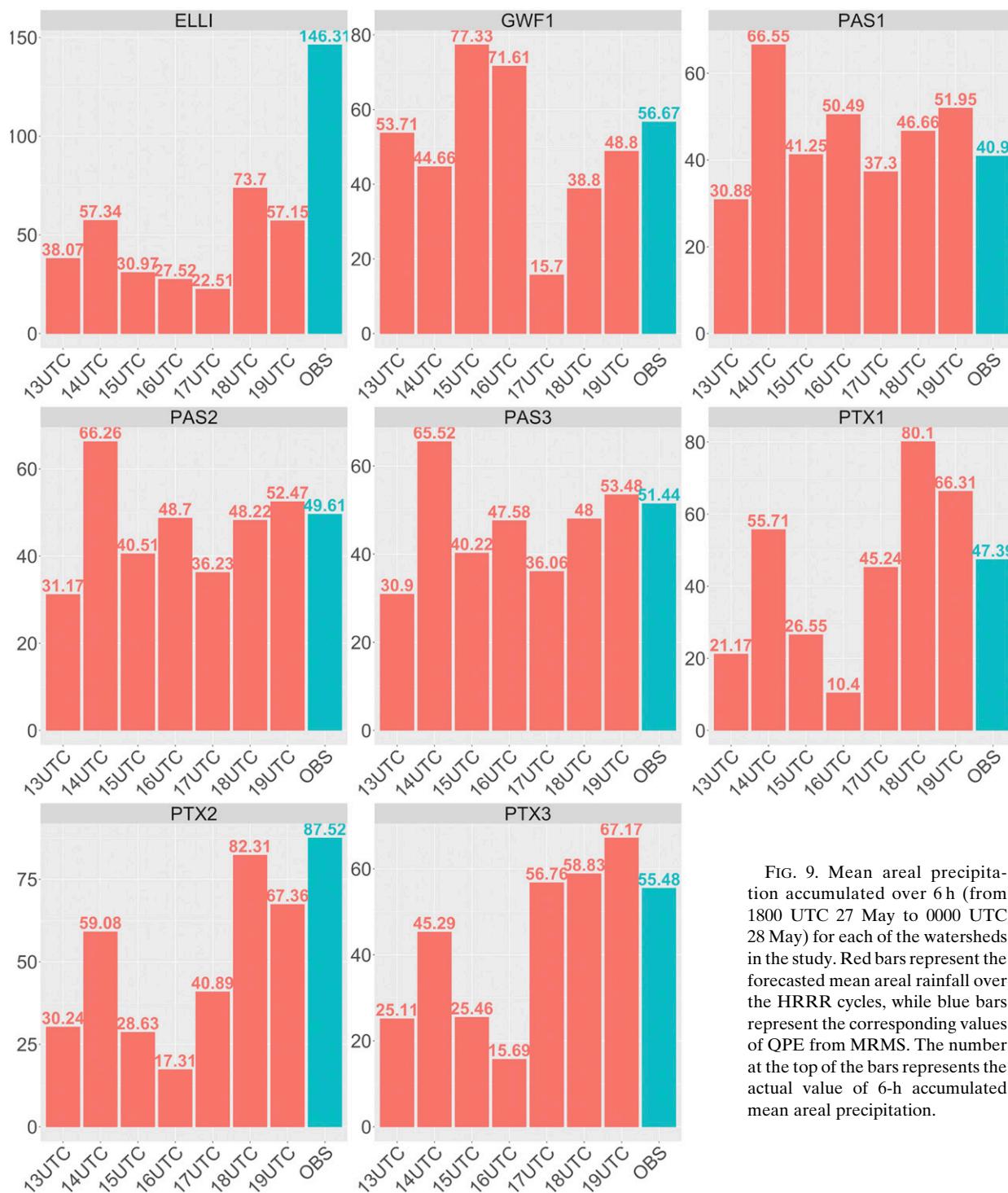


FIG. 9. Mean areal precipitation accumulated over 6 h (from 1800 UTC 27 May to 0000 UTC 28 May) for each of the watersheds in the study. Red bars represent the forecasted mean areal rainfall over the HRRR cycles, while blue bars represent the corresponding values of QPE from MRMS. The number at the top of the bars represents the actual value of 6-h accumulated mean areal precipitation.

several intermediate watersheds (PAS1, PAS2, PTX2) were also evaluated but are not included in the main analysis below since they did not show a significantly different response from the corresponding downstream sections. The NWM analysis cycle is also shown in

Figs. 12 and 13 to both validate against observations and to provide a best estimate of observed flow for the ungauged ELLI watershed. The NWM assimilates gauge observations at all available USGS streamflow locations across its modeling domain, and, consequently, we expect

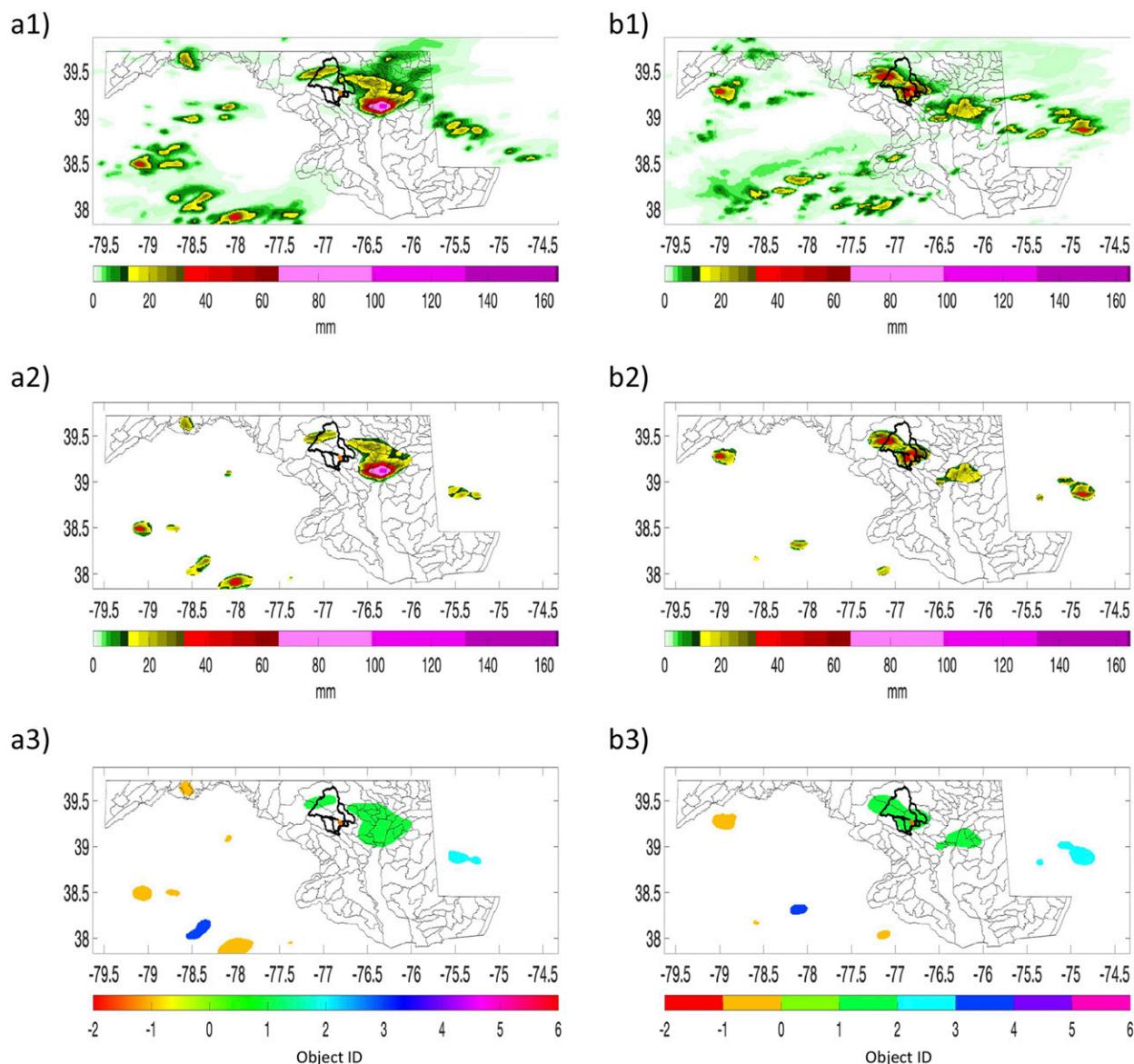


FIG. 10. MET MODE comparison between (left) HRRR 1-h accumulated precipitation from the 1300 UTC cycle and (right) MRMS at 2200 UTC 27 May. (a1) HRRR forecast and (b1) MRMS raw precipitation fields; (a2) HRRR and (b2) MRMS MODE isolated (convoluted and filtered) fields; identified objects from the (a3) HRRR 1300 UTC cycle and (b3) MRMS QPE. In (a3) and (b3), the same “object ID” colors indicates merged (within the same forecast or observation field) and matched MODE objects across forecast and observation fields. The black solid lines in the map represent the watersheds referred to in the text; the ELLI watershed is highlighted in orange.

output from the analysis to match the USGS observations at each USGS site. However, at the ungauged ELLI site, since there are no observations available to assimilate here or upstream, the output is strictly from the model simulation.

Quantitative streamflow evaluation statistics are also summarized in the lower panels of Figs. 12 and 13. Statistical scores used included correlation, NSE, bias, magnitude, and time and volume of the peak to better

evaluate each NWM forecast cycle in the different watersheds.

1) ANALYSIS OF SMALLER WATERSHEDS

As discussed in section 3, Ellicott City was impacted by multiple subhourly pulses of rainfall from 1900 to 2300 UTC 27 May. Due to its hourly time resolution, the MRMS product depicts these subhourly rainfall spells as a single intense rainfall peak from 1900 to 0000 UTC,

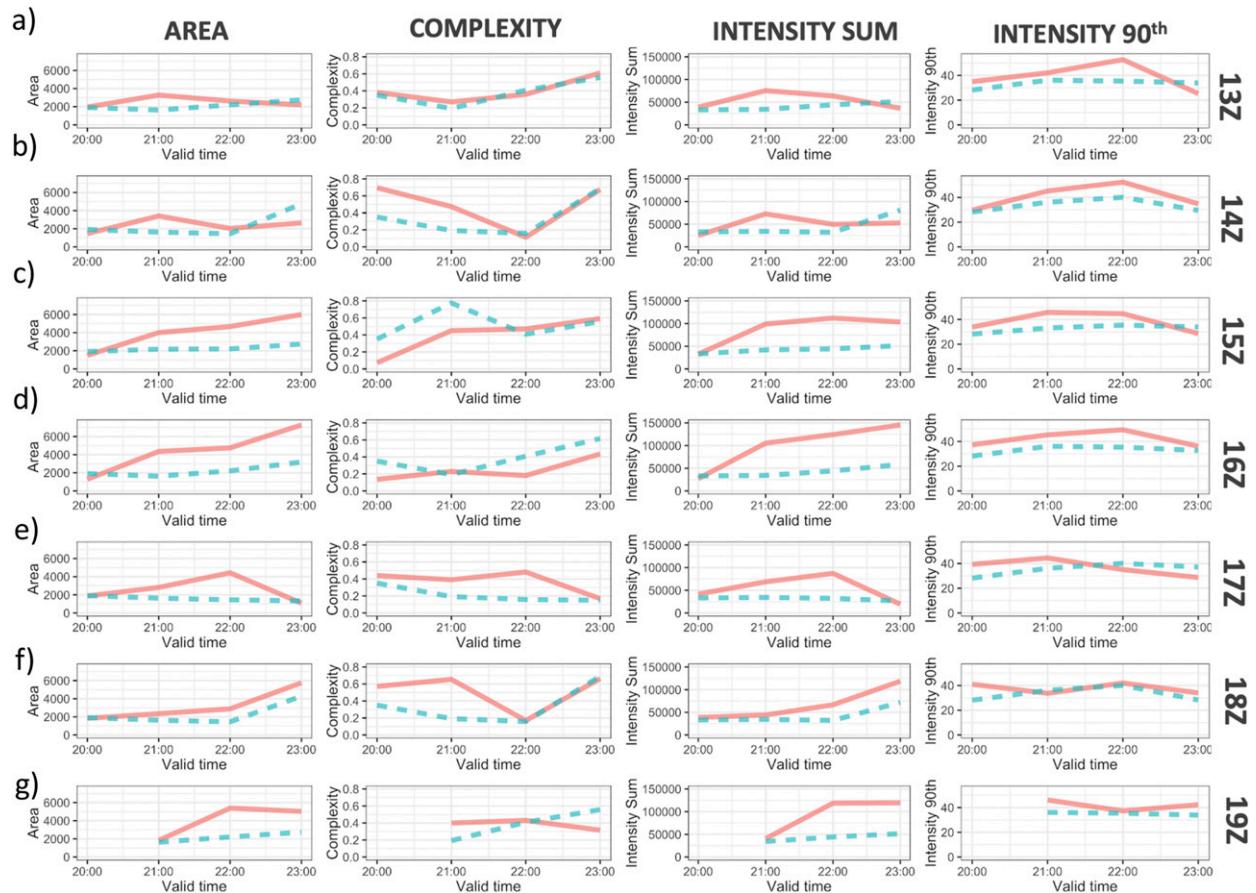
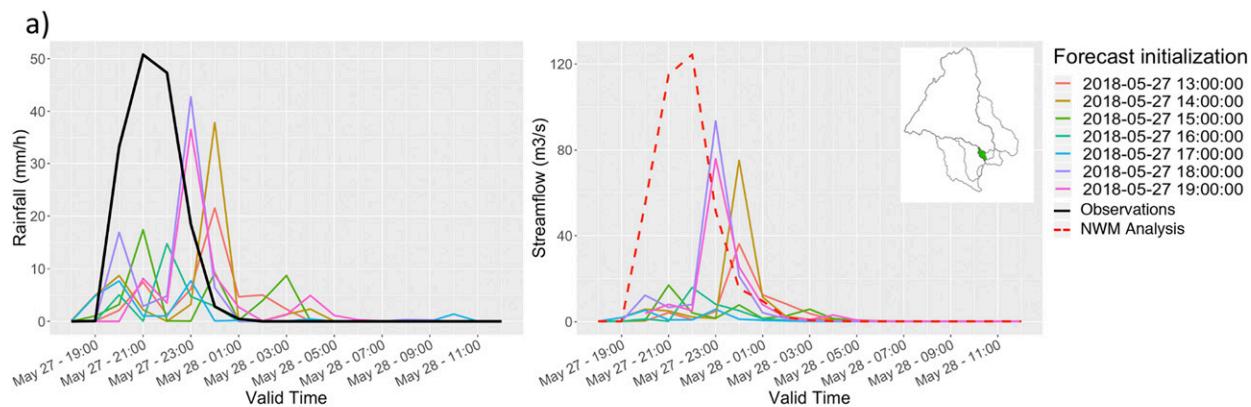


FIG. 11. (a)–(g) Evolution in time of the MODE scores, referred to the object over Ellicott City from MRMS QPE (blue dashed line) and the corresponding paired (merged and matched) one from the HRRR forecast (red solid line). The rows represent the forecast cycles from 1300 UTC to 1900 UTC (from top to bottom), and the columns represent the different scores calculated for the paired objects from MODE (from left to right): area (the area of the object), complexity (the area of the object to the area of its convex hull), intensity sum (the sum of the total intensities within object; mm h^{-1}), and intensity 90th (the 90th percentile of the intensity inside the object; mm h^{-1}).

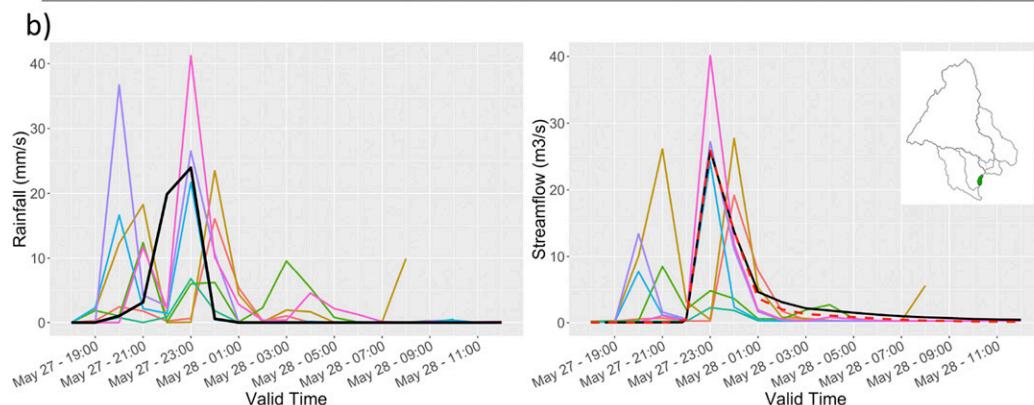
with a maximum at 2100 UTC (Fig. 13a). The HRRR shows some double-peaked rainfall from the 1400, 1800, and 1900 UTC cycles, even though the main rainfall peak is less intense than the MRMS peak and is delayed several hours compared to the MRMS QPE maximum. Over the small ELLI watershed, all other HRRR forecast cycles show either no significant rainfall event or maximum rainfall amounts that are less than half of the MRMS observations.

The NWM streamflow from the short-range cycles for the ungauged ELLI basin closely resembles the rainfall (Fig. 12a): 1900, 1800, and 1400 UTC initializations show the highest streamflow peaks, and most closely match the NWM analyses for streamflow volumes, even if still underestimated. The other less skillful cycles exhibit significantly lower flows and have poor statistics for correlation and time of the peak due to the delayed streamflow response. The NWM streamflow response

for ELLI watershed is thus very strongly linked to the precipitation inputs provided by the NWM meteorological forcings from the HRRR model. Where the HRRR model underestimates precipitation intensity or was delayed in time, the NWM streamflow forecast responds accordingly. The USGS indirect, computed (postmortem) peak streamflow estimates listed in section 3 suggest that the NWM analysis also underestimates flow in the small ELLI basin. Assessing if this is the case is very difficult from these three indirect measurements, since there is significant uncertainty in the methods used to retrieve these values, it is unclear how or whether the maximum peaks may have been coincident, and the hourly time resolution of the NWM does not allow to check if higher maxima were reached within subhourly intervals. Nevertheless, for this watershed, it still appears that the HRRR and the NWM produce forecasts indicating the risk of intense



ELLI Watershed							
	Correlation [-]	NSE [-]	BIAS [m ³ /s]	Peak Flow [m ³ /s]	Time of the Peak [time]	Vol Fcst [m ³]	Vol Obs [m ³]
13z	0.01	-0.23	-16.72	36.14	2018-05-28 00 UTC	48 x 10 ⁵	243 x 10 ⁵
14z	0.02	-0.32	-14.68	74.92	2018-05-28 00 UTC	71 x 10 ⁵	243 x 10 ⁵
15z	0.63	-0.10	-18.33	16.99	2018-05-27 21 UTC	30 x 10 ⁵	243 x 10 ⁵
16z	0.68	-0.11	-18.90	15.91	2018-05-27 22 UTC	23 x 10 ⁵	243 x 10 ⁵
17z	0.37	-0.23	-19.79	5.77	2018-05-27 23 UTC	12 x 10 ⁵	243 x 10 ⁵
18z	0.30	-0.08	-12.53	93.5	2018-05-27 23 UTC	97 x 10 ⁵	243 x 10 ⁵
19z	0.27	-0.08	-13.36	75.65	2018-05-27 23 UTC	87 x 10 ⁵	243 x 10 ⁵
<i>NWM Analysis</i>	-	-	-	124.35	2018-05-27 22 UTC	-	-



PTX1 Watershed							
	Correlation [-]	NSE [-]	BIAS [m ³ /s]	Peak Flow [m ³ /s]	Time of the Peak [time]	Vol Fcst [m ³]	Vol Obs [m ³]
13z	0.40	0.03	-1.08	19.17	2018-05-28 00 UTC	118 x 10 ³	189 x 10 ³
14z	0.20	-1.27	1.60	27.7	2018-05-28 00 UTC	294 x 10 ³	190 x 10 ³
15z	0.43	0.12	-1.45	8.5	2018-05-27 21 UTC	98 x 10 ³	192 x 10 ³
16z	0.97	0.01	-2.57	2.25	2018-05-27 23 UTC	27 x 10 ³	194 x 10 ³
17z	0.85	0.71	-0.84	24.21	2018-05-27 23 UTC	140 x 10 ³	195 x 10 ³
18z	0.88	0.73	0.23	27.2	2018-05-27 23 UTC	211 x 10 ³	196 x 10 ³
19z	0.97	0.70	0.28	40.09	2018-05-27 23 UTC	214 x 10 ³	197 x 10 ³
<i>Obs</i>	-	-	-	25.88	2018-05-27 23 UTC	-	-

FIG. 12. Mean (left) areal precipitation and (right) streamflow for the different forecasts from HRRR and NWM realizations, compared with the observation provided by MRMS for USGS sites (black lines on the left and on the right, respectively). The NWM analysis cycle is represented by the red dashed line as an additional term of comparison for streamflow. Tables under the plots show the statistical performance for the watersheds discussed in the text. The results are reported for watersheds (a) ELLI and (b) PTX1.

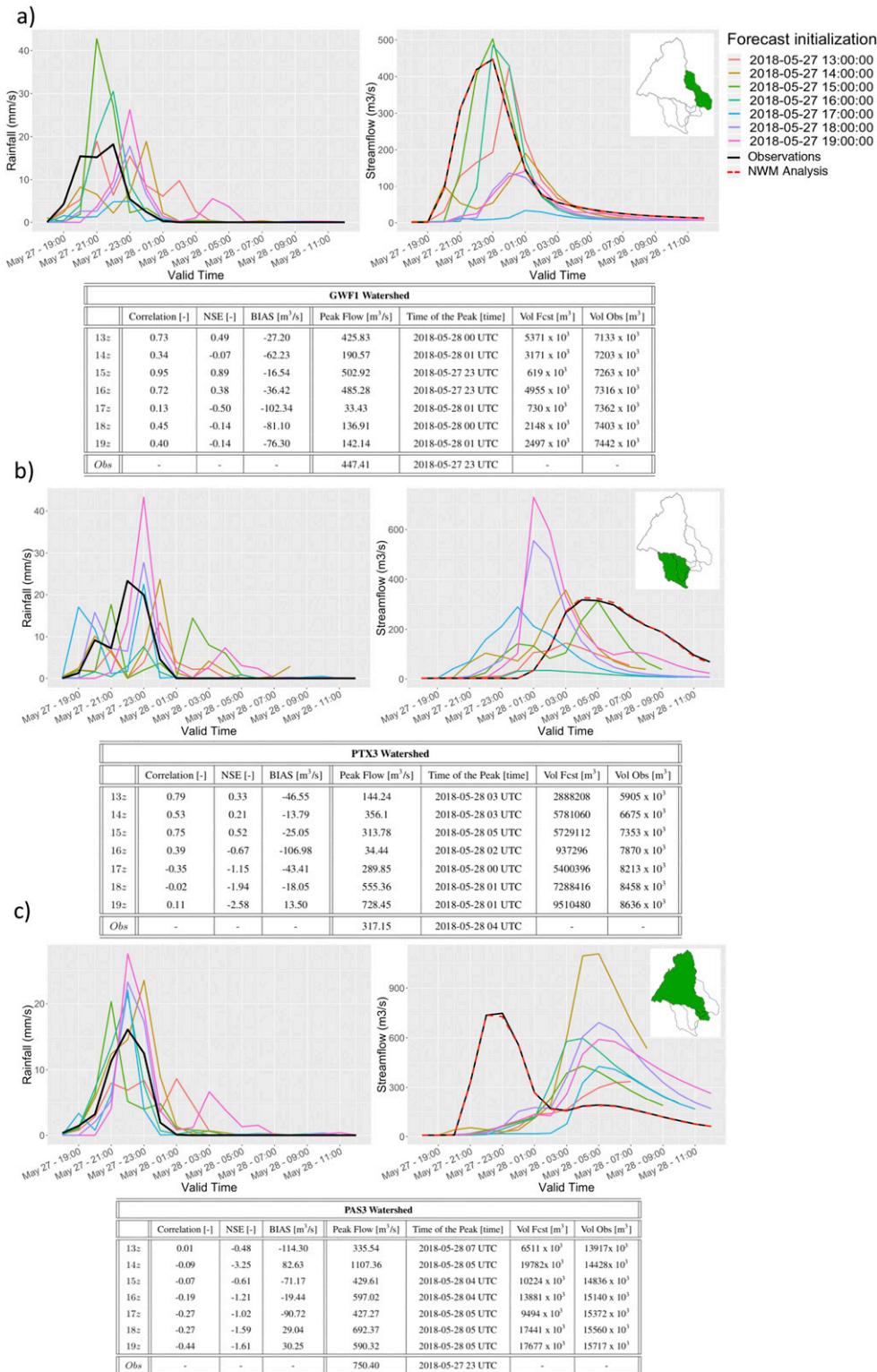


FIG. 13. Mean (left) areal precipitation and (right) streamflow for the different forecasts from HRRR and NWM realizations, compared with the observation provided by MRMS for USGS sites (black lines on the left and on the right, respectively). The NWM analysis cycle is represented by the red dashed line as an additional term of comparison for streamflow. Tables under the plots show the statistical performance for the bigger watershed upstream. The results are reported for watersheds (a) GWF1, (b) PTX3, and (c) PAS3.

rainfall and rapidly rising streamflow in a small, flood-prone basin.

To establish whether the NWM can reasonably reproduce observed streamflow, another watershed of similar size and characteristics to ELLI is evaluated in Fig. 12b. The outlet point of the Little Patuxent watershed (Little Patuxent River tributary above Lake Elkhorn near Guilford, PTX1) was also impacted by a significant flood wave during this event. Figure 12b shows the rainfall and streamflow over this small tributary, in which precipitation was concentrated from 1900 UTC May 27 to 0100 UTC May 28 with the most intense precipitation occurring from 2100 UTC 27 May to 0000 UTC 28 May. Overall, the PTX1 watershed has a very similar behavior to ELLI. The streamflow response is very flashy due to the small catchment size, and the signal of an intense event is clearly represented across the different forecast cycles. Both of the small watersheds shown here indicate that the NWM is very sensitive to this type of meteorological forcing; thus, given relatively skillful HRRR forecasts, the NWM was consequently skillful in indicating potential flood forecast conditions.

2) ANALYSIS OF LARGER WATERSHEDS

Figure 13 shows the rainfall and hydrographs of downstream outlets where a larger upstream area is contributing to the overall streamflow.

GWF1 (Fig. 13a) was first affected by the westward-moving mesoscale storm complex. While some HRRR cycles (1500, 1600, 1900 UTC) significantly overestimate (by nearly double the observed values) the main precipitation peak, this overestimation of precipitation does not translate linearly into an equivalent twofold overestimation of streamflow. Thus, in this watershed NWM streamflow amounts do not appear to directly correspond to given rainfall volumes. Here, the NWM appears to underproduce runoff, leading to proportional floodwave representation. The analysis of this specific case suggests further investigation may be useful for multiple events or longer time series, as it seems to point out that changes to NWM runoff generation mechanisms or relevant model parameters may be beneficial.

Over the Patuxent (PTX3) (Fig. 13b) the HRRR precipitation forecasts possess some variability in time and space, with both high and low precipitation biases depending on the specific cycle. While the rainfall forecasts show timing errors in both directions (too early and too late), the NWM streamflow forecasts are systematically too early. Thus, the watershed response for this basin appears to be faster in the NWM than what is actually observed.

Finally, PAS3 represents how the event evolved along the main stem of the Patapsco [which partially includes

the upstream contribution of the Tiber River in Ellicott City (ELLI)], and thus also depicts forecast flood response in one of the main flood-impacted regions. Figure 13c shows observed peak precipitation between 1900 UTC 27 May and 0100 UTC 28 May with a maximum peak reached at 2200 UTC. Observed streamflow shows a corresponding double-peak where the first intense peak ($750.4\text{ m}^3\text{ s}^{-1}$) is reached at 2200 UTC, while a second, much less intense hydrograph rise is registered from 0200 UTC to 0700 UTC 28 May. The NWM streamflow forecast realizations exhibit a significantly different shape from the observed hydrograph here: the initial rising limb is significantly underestimated or completely missing in all NWM forecasts, while the second peak is overestimated or, possibly, quite delayed. Because of the aforementioned importance of the PAS3 section, additional analysis is performed below.

3) PATAPSCO RIVER ANALYSIS

The discordance between observed and forecasted streamflow in the PAS3 section motivated additional analysis over the Patapsco River. Here, additional river sections have been considered along the main stem of the Patapsco in order to understand the relationship between flood wave propagation signals and NWM forcings both upstream and downstream of the confluence in Ellicott City. We specifically highlight the PAS2 section, which is immediately downstream of the confluence in Ellicott City, and also use another available USGS streamflow gauge just upstream of Ellicott City (USGS 01589000), as illustrated in the map of Fig. 14.

First, to understand the observed hydrograph evolution, and in particular flood wave propagation, Figs. 14a and 14c shows that the flood wave propagation time from USGS 01589000 (upstream of Ellicott City) to the PAS2 outlet (downstream of Ellicott City) is approximately 1 h (blue dotted line). Further, the observed shape of the hydrograph upstream of Ellicott City (black line in Fig. 14a) exhibits a single peak at 2100 UTC, while observed streamflow in PAS2 (black line in Fig. 14c) has a double-peaked shape similar to that of the PAS3 section (shown in Fig. 13c). With respect to timing, the red, dashed vertical line in Figs. 14b and 14c shows the close correspondence in time (i.e., less than 1 h) between the main peak occurring in Ellicott City and the one observed in PAS2. This suggests that the contribution of the first high-peak in the observed hydrographs of PAS2 (at 2100 UTC) and PAS3 (1 h later) comes from the contribution of stream volume from Ellicott City floodwaters, while the second later and smaller observed rise (at 0400 UTC) of the hydrographs

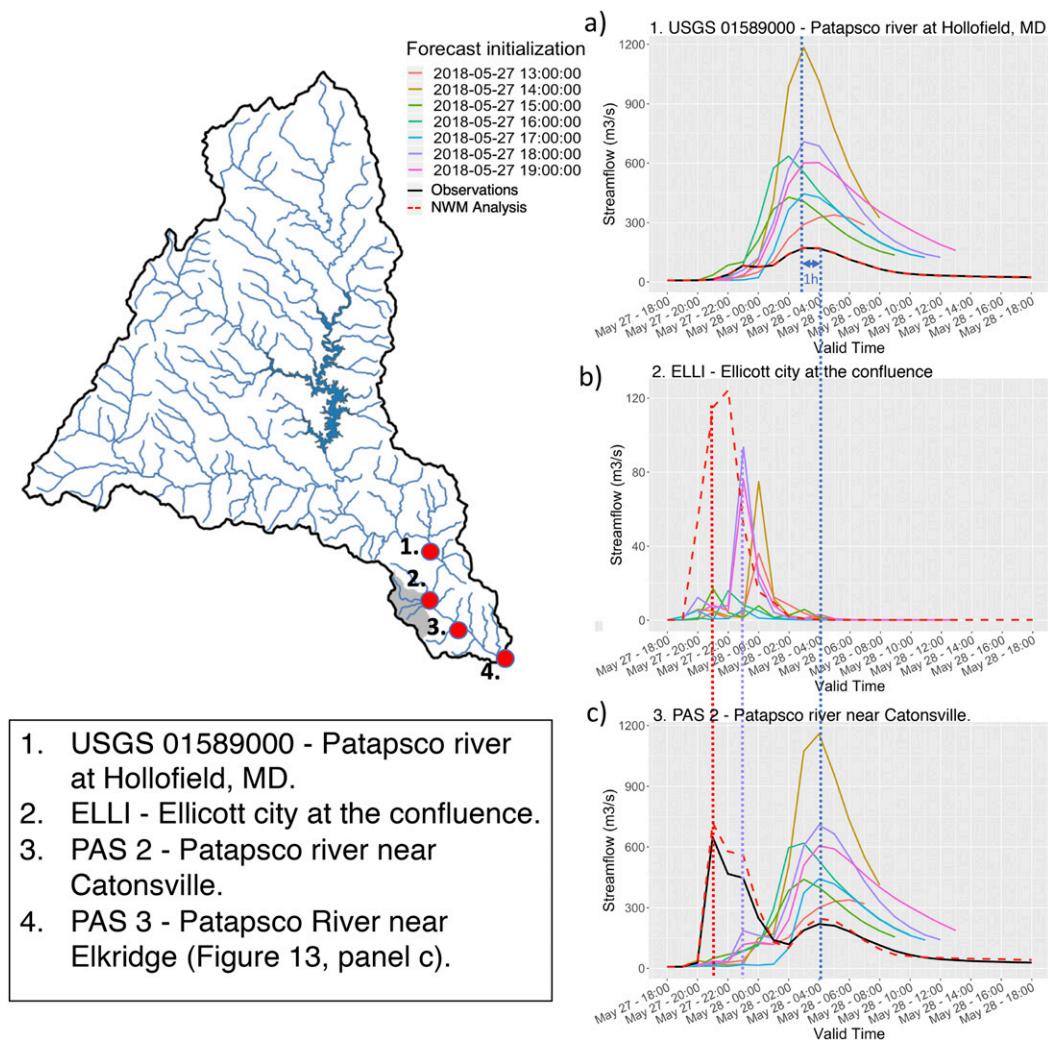


FIG. 14. Additional sections considered for the flood propagation analysis along the Patapsco, with related hydrographs. (left) A map of the Patapsco watershed and the sections examined are reported in the legend. (right) Hydrographs for the (a) USGS 01589000 section, (b) ELLI section, and (c) PAS2 section are represented. Vertical dotted lines in the figure offer a reference to the reader to interpret the related text. The red dotted line represents the correspondence between the peak happening in Ellicott City in (b) and PAS2 in (c). The violet line represents the peak in the 1800 UTC cycle for ELLI in (b) and PAS2 in (c). The blue dotted line connects the hydrograph rise at 0400 UTC from USGS 01589000 to the PAS2 section peak 1 h later, to indicate the travel time of the flood propagation wave between these two sections.

derive more from runoff contributions coming from upstream.

The evolution of the NWM streamflow forecasts upstream and downstream of Ellicott City differs significantly from observations. The first observed peak is notably underforecast for both PAS2 and PAS3 (Figs. 14c, 13c). This is likely due to delayed and underforecast HRRR precipitation in the ELLI basin (Fig. 12a). As discussed above, from a forecaster situational awareness perspective, the QPF in this basin was sufficient to produce streamflow indicative of a potential flood response during the event in general. However, in

terms of hydrologic impacts, here we see that critical details in precipitation forcing (too little and too late) and (possibly) hydrological model errors over this small basin critically compound downstream forecast errors found at PAS2 and PAS3. With respect to the second observed hydrograph peak (0400 UTC for PAS2; Fig. 14c), the NWM overestimation here is likely largely explained by significant QPF overestimation in the far (northwest) upstream reaches of PAS3 (Fig. 15). As these upstream errors translate downstream, they manifest in continued streamflow overestimation downstream of Ellicott City (Fig. 14c).

Thus, the competing effects of overdone QPF far upstream in PAS3 and underdone, delayed QPF in the ELLI basin result in noteworthy streamflow errors depicted in Fig. 13c. We use this in-depth analysis on the Patapsco to demonstrate the complex error interactions that result from combined meteorological and hydrologic uncertainties. Taking a careful, multiscale, hydrometeorological approach is essential to understand hydrologic model behavior, and an appreciation of all potential NWM error sources is critical for both future model development, as well as for appropriate and informed use by forecasters.

4) DISCUSSION OF NWM STREAMFLOW EVALUATION

The streamflow analysis presented highlights salient forecast challenges that can vary with basin size. Over smaller basins affected by this particular flood event (ELLI and PTX1), the NWM response to generally skillful HRRR QPF appears to be both physically reasonable and accordingly skillful. Such capabilities highlight that, particularly in small, ungauged locations, provided skillful precipitation forecast forcing, NWM analysis data and forecasts may offer useful flash flood guidance (an especially useful potential complement to operational NWS streamflow forecasts in areas that do not currently receive them).

For this event, when NWM forecasts are evaluated over larger basins, lower NWM forecast skill is found, along with increased complexity in explaining the errors. Over larger areas, additional factors must be considered for these watersheds; for example, the distance between the most intense rainfall and the watershed outlet point, increased heterogeneity in land use and soil type, as well as the contribution of propagating errors from other contributing basins [as demonstrated in section 5b(3)]. In addition to that, because the water must travel across longer distances in larger basins, sources of meteorological and hydrological uncertainties can combine and grow quickly upscale, rendering even small displacements of precipitation in time and space more likely to significantly affect the final streamflow performance. Note that this particular result is likely specific to the examination of individual events. More traditional hydrologic forecast assessments over longer periods are more likely to show enhanced skill for larger basins, as small-scale errors average (effectively self-canceling) out over time.

Finally, as stated by many studies (Bytheway et al. 2019, and references therein, Nelson et al. 2016; Smalley et al. 2014), possible uncertainties in the radar based and gauge-corrected QPE can always contribute to additional sources of uncertainties in the evaluation.

While completely isolating the causes of streamflow forecast errors is well beyond the scope of this study (and remains a grand challenge in hydrometeorological science), some additional investigation and sensitivity tests were performed to consider: (i) possible impacts of land use change in the NWM, (ii) effects of streamflow observation assimilation, (iii) sensitivity to antecedent soil moisture conditions, (iv) surface and subsurface runoff versus baseflow partitioning, and (v) NWM model version. Given the aforementioned complexities involved in disentangling the role of forcings, it is not surprising that no clear singular error source emerged from these analyses, but as they present possibilities for future research, we briefly summarize the exploration of these effects:

- (i) The land use analysis comparing the NLCD 2011 land use (as used in the NWM version 1.2) with the most recent NLCD 2016 product (Yang et al. 2018) showed that the change in urban land cover was less than 13%, and thus likely not significant enough to critically affect the possible differences in soil infiltration capabilities and runoff (Wigmosta and Burges 2001).
- (ii) Investigation of the impact of streamflow observation assimilation demonstrated that without assimilation, the NWM tends to overestimate streamflow in the Upper Patapsco, but that the magnitude of this effect is quite small compared to the HRRR overestimation at that time and location, again underscoring the difficulty in quantifying the relative impact of all individual error sources.
- (iii) The role of antecedent soil moisture in flash flood generation was considered despite the lack of soil moisture observations in the area. The operational NWS tool known as “flash flood guidance” (FFG) was used to indicate relative soil saturation conditions, integrating analyzed current soil moisture and streamflow conditions to estimate the amount of rainfall required for a given time and area to produce bankfull conditions on small streams (Schmidt et al. 2007; Clark et al. 2014). Comparing the operational FFG, HRRR QPF, and NWM initial condition soil infiltration capacity demonstrated that the FFG and NWM initial condition are very similar, and each are exceeded by HRRR QPF on hourly and 6-hourly time scales. Therefore, anomalous NWM soil moisture conditions do not appear to be a significant NWM error source for this event, as might be expected for an intense, localized infiltration excess type of flash flood event in a watershed with the shallow soil and steep terrain as in Ellicott City.

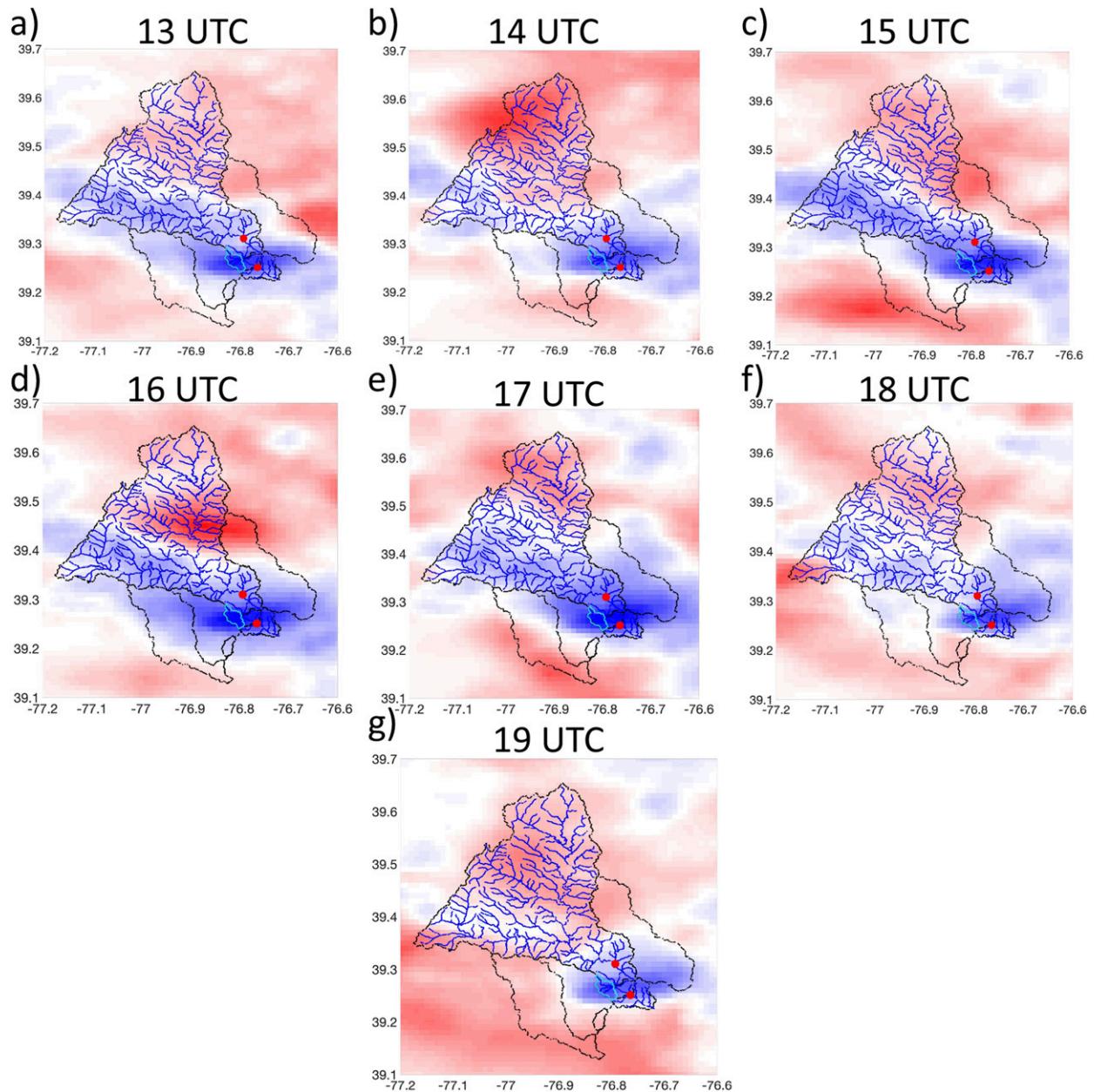


FIG. 15. (a)–(g) Accumulated bias between HRRR cycles initialized from 1300 to 1900 UTC 27 May 2018 and MRMS QPE, as obtained from the full 18-h duration of the short-range cycle using the Series Analysis tool from MET. The blue color means underestimation of the HRRR cycle, as compared to the MRMS, while the red color means overestimation of the HRRR cycle over the MRMS QPE. The red dot upstream represents USGS 01589000, and the red dot downstream represents the PAS2 section. The light blue contour indicates the Ellicott City watershed.

- (iv) Relative contributions of surface flow and base flow showed practically zero contributions from the baseflow bucket model, confirming that the NWM was likely appropriately partitioning surface and subsurface runoff for this event.
- (v) Finally, the potential for forecast improvement due to changes made in subsequent NWM versions was

also explored, in particular comparing the performance of the model at the time of this event (version 1.2) with the subsequent operational version 2.0. NWM version 2.0 introduced the out-of-bank parameterization via compound channel and new empirically based parameters that, together with improved and extended calibration, might

affect the NWM results for the Ellicott City event. While the newer model version yielded some minor improvements for some of the streamflow sections, it did not solve the salient behaviors discussed above.

Though no single hypothesis test revealed a primary NWM error source, all of these possibilities are worth exploring more deeply in future work. With sufficient additional cases and longer-term evaluation, such exercises should expose and explain specific model error sources and possible routing limitations such that results can directly benefit ongoing model development and make lasting improvements to future model versions.

c. Inundation areas

As noted in section 4b(1), the NWM outputs several distributed hydrological variables, but it does not directly calculate flood inundation extent or depth. Methods exist to create flood inundation maps from the NWM output and provide potential benefit to stakeholders (NOAA 2015). While these methods are currently experimental, NOAA is working with other federal partners to eventually bring a flood inundation mapping capability into operations. Following the method described in section 4c, Fig. 16 shows the maximum extent of the flood inundation areas from forecast cycles initialized from 1300 to 1900 UTC 27 May, compared against the same inundation calculated from the NWM Analysis (Fig. 16, red bold line). Given that the NWM analysis seems to be a reasonable surrogate for ground truth in the small ELLI region hit hardest by flash flooding (as shown in the ELLI streamflow evaluation), the flood inundation maps produced from the NWM streamflow and the HAND method extend the potential utility of the NWM outputs by providing a spatial distribution of surface water and thus another perspective of the potential impact of the flood over the urban area.

Specifically, the 1400, 1800, and 1900 UTC cycles (Fig. 16) closely match the flood inundation area, although producing some overestimation over the Hudson branch area, north of Main Street. In contrast, the 1300, 1500, 1600, and 1700 UTC cycles tend to create less flooding than the flood maximum flood from the NWM analysis. In particular Fig. 16 highlights the 1800 UTC cycle (Fig. 16, filled azure contour) and the 1700 UTC cycle (Fig. 16, light aqua filled contour), as they are the forecast cycles that better matches the NWM Analysis and the ones that produce the smallest inundation extents, respectively. The smallest flood inundation extent of the 1700 UTC cycle is consistent with the largest underestimation of mean areal precipitation (Fig. 9) and the lowest forecasted discharge (Fig. 12a) in the ELLI

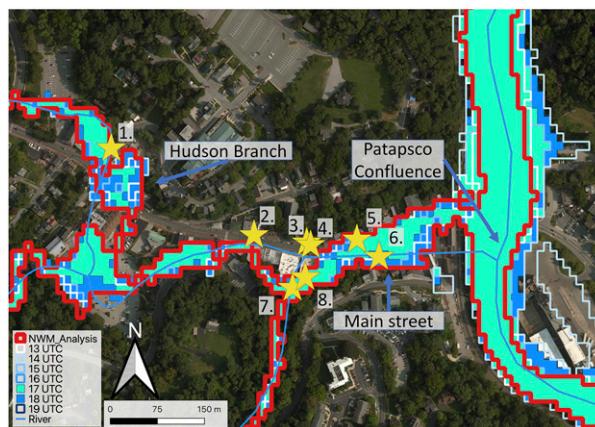


FIG. 16. Maximum extent of the flood inundation areas from forecast cycles initialized from 1300 (white line) to 1900 UTC (cobalt line) 27 May, compared against the same inundation calculated from the NWM analysis (bold red line).

watershed. Overall, however, all the forecast cycles are consistent with the NWM analysis, reaching maximum inundation extent in similar areas.

One major challenge in the development of flood inundation maps is that observations of urban flooding are uncommon. Here, we use a unique collection of 12 street cameras that were installed following the 2016 flood (Ellicott City Unilux Camera Network; Peters 2019) in order to compare the temporal evolution of the inundation map indicated from 2100 UTC 27 May to 0000 UTC 28 May. The temporal evolution analysis is performed only for the 1800 UTC forecast cycle (as it best matched observations both in terms of rainfall and streamflow and therefore serves as a useful prototype) (Fig. 17). Finally, we note that the areas surrounding the Hudson Branch near Rogers Avenue are not included in the NWM network due to NHDplus limitations, thus the Roger Main camera location (number 9 in Fig. 6) is discarded from the analysis.

The inundation extent maps produced from the 1800 UTC NWM forecast cycle show remarkable agreement with street camera observations. In particular, with respect to timing, the street cameras show maximum inundation extent occurring around 2300 UTC, and the forecast flood inundation maps reflect this timing as well. Furthermore, the cameras capture floodwaters beginning to recede between 2300 UTC and 0000 UTC, and this recession is also indicated by the flood inundation maps at 0000 UTC. Similar behavior in terms of maximum and regression is observed also for the other street camera locations, although limitations are found in locations confined by buildings that are spaced close together (e.g., the Tiber Alley camera is located in a very small restaurant courtyard).

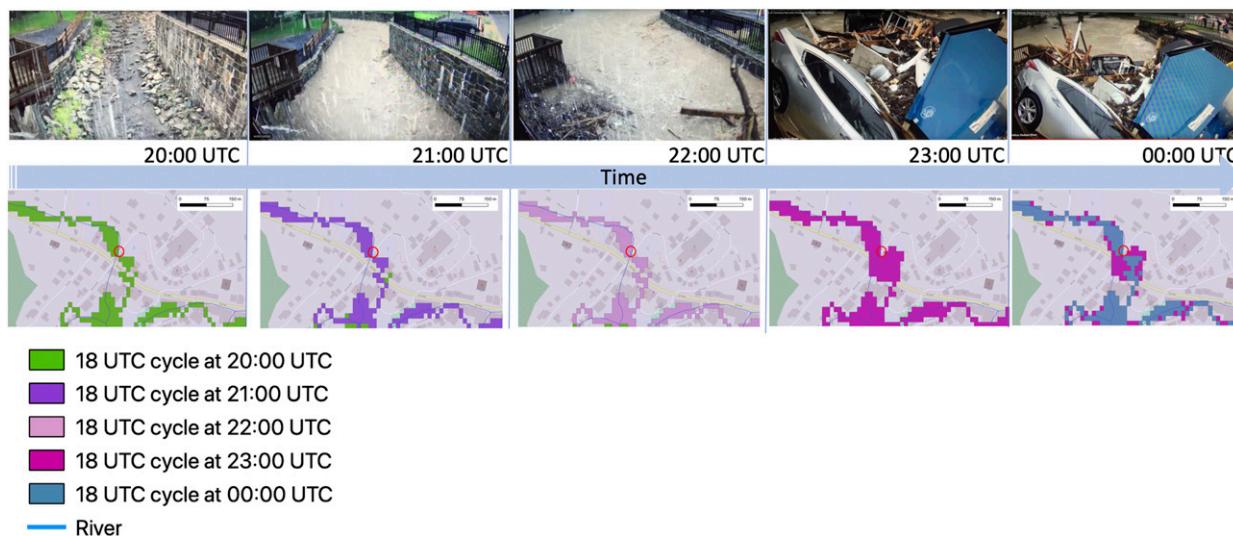


FIG. 17. Hourly evolution of the forecast mapped flood extent from 2100 UTC 27 May to 0000 UTC 28 May from the forecast cycle initialized at 1800 UTC. The red dot represents the camera footage and orientation.

While the NWM does not currently provide comprehensive inundation mapping as a native model output, it is an area of ongoing work, and based on the evaluation demonstrated here, the potential for forecast guidance of surface flood extent and timing hours in advance could be of significant help to forecasters, as well as preparing for urban flood impacts and response such as evacuation, preemptive road closure, and a potential new angle to risk communication.

6. Conclusions

This multiscale evaluation of National Water Model short-range forecasts during a high-impact flash flood event in Ellicott City, Maryland, in May 2018 demonstrates both the advantages and challenges of using the NWM, a CONUS-scale, distributed hydrologic modeling system, to accurately predict the impacts of a local-scale, short-duration flood event. The analysis proceeds as a series of verification approaches across spatial scales, beginning with the verification of HRRR model precipitation forecasts as inputs to the NWM, then moving to regional watershed-scale evaluations, and finally to the urban flooding scale in which experimental flood inundation maps are evaluated using uniquely available footage from street cameras.

While this particular study examines a single event, it establishes a process for a joint hydrometeorological evaluation method which is critical to understanding—and ultimately improving—hydrologic model inputs and outputs. We use local impacts and observations of extreme events to begin disentangling multiple error and skill sources from a complex modeling and forecast

system. Specifically for this event, mesoscale QPF verification reveals that the HRRR model produced skillful forecast guidance for a high-intensity mesoscale event over the Maryland region. Nevertheless, small displacements on the order of a just few kilometers in space and a few hours in time in the atmospheric inputs can strongly influence the hydrological response at the watershed scale, and shifts in cycle-to-cycle HRRR forecast skill are accordingly seen in the resulting NWM skill as well. Complementary gridded and object-based precipitation verification approaches further demonstrate the importance of a comprehensive approach to QPF performance evaluation, especially when event impacts are so sensitive to both atmospheric and hydrological uncertainties. For example, in the forecast cycles investigated here, gridded verification suggests that HRRR QPF was underestimated over the Ellicott City region, while the object-based verification demonstrates that HRRR correctly represented both precipitation intensity and the salient characteristics of the mesoscale storm, but was spatially offset. The hydrologic ramifications of the spatial shifts from an otherwise skillful QPF highlights future opportunities in applying probabilistic approaches, even using deterministic model output.

Watershed scale evaluation of the NWM suggests that the hydrological response from the NWM is more complex in larger basins, where the propagation and accumulation of uncertainties from the meteorological forcings to the streamflow outputs influence the resulting hydrologic forecasts more profoundly. Additional investigation at still finer scales is required to more completely explore the hypotheses posed here regarding specific NWM error sources. In addition to streamflow

verification at the watershed scale, the potential utility of using the NWM in ungauged areas is also demonstrated. For small-scale watersheds, the NWM response is linked very clearly to the meteorological forcing inputs. Intense rainfall directly translated to an intense flooding event in both of the smaller watersheds examined. While errors in timing and volumes at these scales can be attributed to both the quality of forcings and hydrologic model response, across consecutive model forecast cycles, indications of the possibility of a fast-onset flash flood event would likely have offered short-term forecast utility.

Finally, at the urban scale, we demonstrate that experimental HAND method-based flood inundation maps may offer promise for anticipating the extent and the timing of the flood inundation areas. Because of its computational simplicity, the HAND method, when coupled with the NWM streamflow forecasts, can quickly yield flood inundation maps. As such, the coupling of HAND and NWM may be a valuable operational forecasting asset for emergency weather preparedness and response, where high forecast temporal resolution is often preferred. For the specific case of the Ellicott City event, maps compared to street video camera footage reveal the potential benefit and skill of this NWM-driven capability. While this method remains under ongoing development, a successful forecast example is shown here in the accurate prediction of urban flood inundation with 2-h lead time.

In summary, the evaluation of the NWM for the May 2018 Ellicott City, Maryland, flood suggests potential advantages and forecast utility in using the CONUS-scale, high-resolution, distributed NWM to predict high-impact, local scale flood events, while also underscoring the need to comprehensively evaluate model performance at local scales and for high-impact, rare events. A salient challenge identified by the combination of hydrologic and meteorological evaluation is that while streamflow is the standard, and in most ways most straightforward, hydrologic model output field to verify, it is merely a single output of a modeling system that is simultaneously representing and predicting many complex processes above and below Earth's surface. Specifically, model representation of soil infiltration capacity and soil saturation, heat and moisture fluxes on land and runoff partitioning, heterogeneity of the geographical domain, and model calibration are challenges at and below the surface; meteorological forcing (precipitation in particular) errors represent an additional, and often primary, source of uncertainty to successful hydrometeorological prediction. This case study also clearly demonstrates that even with generally skillful precipitation forecast inputs at the larger

mesoscale, there are yet additional hydrologic basin-scale sensitivities that highlight a dilemma: hydrologic response will be simpler and likely more skillful over a small, limited-area basin. However, precise and accurate precipitation forecasts over small areas (particularly for heavy convective rainfall flash flood events) remain an outstanding atmospheric NWP challenge. At larger scales, QPF may be considered more predictable, but complex hydrometeorological interactions accumulate errors. Ongoing regional and case study analysis is necessary to more fully understand both the benefits and risks of using a national hydrologic model tool, such as the NWM, to connect CONUS-scale forcings to local scale impact predictions. Through such future work the larger hydrometeorological forecast community can better understand the potential of this model and others like it to give hydrological forecast guidance and to optimally complement existing operational local models to provide the most useful forecast guidance across scales of interest.

Acknowledgments. This study was supported by the Physical Sciences Division of the NOAA Earth System Research Laboratory. We would like to acknowledge high-performance computing support from Cheyenne ([doi:10.5065/D6RX99HX](https://doi.org/10.5065/D6RX99HX)) provided by NCAR's Computational and Information Systems Laboratory, sponsored by the National Science Foundation. Finally, we thank the Ellicott City Unilux Camera Network ([Peters 2019](https://doi.org/10.5065/D6RX99HX)) for the observation provided for the flood inundation map validation.

REFERENCES

- Benjamin, S. G., and Coauthors, 2016: A North American hourly assimilation and model forecast cycle: The Rapid Refresh. *Mon. Wea. Rev.*, **144**, 1669–1694, <https://doi.org/10.1175/MWR-D-15-0242.1>.
- Brown, B. G., J. H. Gotway, R. Bullock, E. Gilleland, T. Fowler, D. Ahijevych, and T. Jensen, 2009: The model evaluation tools (MET): community tools for forecast evaluation. *25th Conf. on International Interactive Information and Processing Systems (IIPS) for Meteorology, Oceanography, and Hydrology*, Phoenix, AZ, Amer. Meteor. Soc., 9A.6, <http://ams.confex.com/ams/pdfpapers/151349.pdf>.
- Bullock, R., B. Brown, and T. Fowler, 2016: Method for object-based diagnostic evaluation. NCAR Tech. Note NCAR/TN-532+STR, 84 pp., <https://doi.org/10.5065/D61V5CBS>.
- Bytheway, J. L., M. Hughes, K. Mahoney, and R. Cifelli, 2019: A multiscale evaluation of multisensor quantitative precipitation estimates in the Russian River basin. *J. Hydrometeorol.*, **20**, 447–466, <https://doi.org/10.1175/JHM-D-18-0142.1>.
- Clark, R. A., J. J. Gourley, Z. L. Flamig, Y. Hong, and E. Clark, 2014: CONUS-wide evaluation of National Weather Service flash flood guidance products. *Wea. Forecasting*, **29**, 377–392, <https://doi.org/10.1175/WAF-D-12-00124.1>.

- Collier, C., 2007: Flash flood forecasting: What are the limits of predictability? *Quart. J. Roy. Meteor. Soc.*, **133**, 3–23, <https://doi.org/10.1002/QJ.29>.
- Davis, C., B. Brown, and R. Bullock, 2006: Object-based verification of precipitation forecasts. Part II: Application to convective rain systems. *Mon. Wea. Rev.*, **134**, 1785–1795, <https://doi.org/10.1175/MWR3146.1>.
- Developmental Testbed Center, 2018: MET: Version 8.0 model evaluation tools users guide. Accessed 8 February 2019, <http://www.dtcenter.org/met/users/docs/overview.php>.
- Garbrecht, J., and G. W. Brunner, 1991: A Muskingum-Cunge channel flow routing method for drainage networks. USACE Tech. Paper 135, 35 pp., <https://www.hec.usace.army.mil/publications/TechnicalPapers/TP-135.pdf>.
- Gochis, D., W. Yu, and D. Yates, 2015: The WRF-Hydro model technical description and user's guide, version 3.0. NCAR Tech. Doc., 120 pp., http://www.ral.ucar.edu/projects/wrf_hydro/.
- Gourley, J. J., and Coauthors, 2017: The FLASH Project: Improving the tools for flash flood monitoring and prediction across the United States. *Bull. Amer. Meteor. Soc.*, **98**, 361–372, <https://doi.org/10.1175/BAMS-D-15-00247.1>.
- Hazewinkel, M., 2013: *Encyclopaedia of Mathematics: C an Updated and Annotated Translation of the Soviet "Mathematical Encyclopaedia"*. Vol. 2, Springer Science & Business Media, 508 pp.
- Homer, C., and Coauthors, 2015: Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information. *Photogramm. Eng. Remote Sens.*, **81**, 345–354.
- Julien, P. Y., B. Saghafian, and F. L. Ogdén, 1995: Raster-based hydrologic modeling of spatially-varied surface runoff. *J. Amer. Water Resour. Assoc.*, **31**, 523–536, <https://doi.org/10.1111/j.1752-1688.1995.tb04039.x>.
- Liu, Y. Y., D. R. Maidment, D. G. Tarboton, X. Zheng, and S. Wang, 2018: A CyberGIS integration and computation framework for high-resolution continental-scale flood inundation mapping. *J. Amer. Water Resour. Assoc.*, **54**, 770–784, <https://doi.org/10.1111/1752-1688.12660>.
- Maryland Department of Natural Resources Watershed Services, 2005: Characterization of the Patapsco River lower north branch watershed in Howard County, Maryland. 64 pp., https://dnr.maryland.gov/waters/Documents/WRAS/patlnb_char.pdf.
- McKay, L., T. Bondelid, T. Dewald, J. Johnston, R. Moore, and A. Rea, 2012: NHDPlus Version 2: User Guide. U.S. EPA Tech. Doc., 182 pp., ftp://ftp.horizon-systems.com/nhdplus/NHDPlusV2/Documentation/NHDPlusV2_User_Guide.pdf.
- Mockus, V., 1964: *National Engineering Handbook*. U.S. Soil Conservation Service, 548 pp.
- Nash, J. E., and J. V. Sutcliffe, 1970: River flow forecasting through conceptual models part I: A discussion of principles. *J. Hydrol.*, **10**, 282–290, [https://doi.org/10.1016/0022-1694\(70\)90255-6](https://doi.org/10.1016/0022-1694(70)90255-6).
- National Weather Service, 2018: Ellicott City & Catonsville, Maryland - Heavy Rain and Flash Flooding of May 27th, 2018. Accessed 8 February 2019, <https://www.weather.gov/lwx/EllicottCityFlood2018>.
- Nelson, B. R., O. P. Prat, D.-J. Seo, and E. Habib, 2016: Assessment and implications of NCEP Stage IV quantitative precipitation estimates for product intercomparisons. *Wea. Forecasting*, **31**, 371–394, <https://doi.org/10.1175/WAF-D-14-00112.1>.
- Niu, G.-Y., and Coauthors, 2011: The community Noah land surface model with multiparameterization options (Noah-MP): 1. Model description and evaluation with local-scale measurements. *J. Geophys. Res.*, **116**, D12109, <https://doi.org/10.1029/2010JD015139>.
- NOAA, 2015: Flash flood services for the future: Flash Flood Summit and Focus Group findings. 74 pp., https://www.weather.gov/media/water/NOAA_Flash_Flood_Summit_Report5_29_15.pdf.
- Nobre, A. D., L. A. Cuartas, M. R. Momo, D. L. Severo, A. Pinheiro, and C. A. Nobre, 2016: HAND contour: A new proxy predictor of inundation extent. *Hydrol. Processes*, **30**, 320–333, <https://doi.org/10.1002/hyp.10581>.
- Office of Water Prediction, 2016: The National Water Model. Accessed 8 February 2019, <https://water.noaa.gov/about/nwm>.
- Ogden, F. L., 1998: CASC2D version 1.18 reference manual. USACE Rep., 106 pp.
- Peters, R., 2019: The Ellicott City, Maryland, floods of 2016 and 2018: Weather-Ready Nation perspectives from a small business owner and landlord. *Seventh Symp. on Building a Weather-Ready Nation*, Phoenix, AZ, Amer. Meteor. Soc., 1.5, <https://ams.confex.com/ams/2019Annual/webprogram/Paper354081.html>.
- Rennó, C. D., A. D. Nobre, L. A. Cuartas, J. V. Soares, M. G. Hodnett, J. Tomasella, and M. J. Waterloo, 2008: HAND, a new terrain descriptor using SRTM-DEM: Mapping terra-firme rainforest environments in Amazonia. *Remote Sens. Environ.*, **112**, 3469–3481, <https://doi.org/10.1016/j.rse.2008.03.018>.
- Rodda, H. J., 2005: The development and application of a flood risk model for the Czech Republic. *Nat. Hazards*, **36**, 207–220, <https://doi.org/10.1007/s11069-004-4549-4>.
- Rossa, A., P. Nurmi, and E. Ebert, 2008: Overview of methods for the verification of quantitative precipitation forecasts. *Precipitation: Advances in Measurement, Estimation and Prediction*, S. C. Michaelides, Ed., Springer, 419–452.
- Schmidt, J. A., A. Anderson, and J. Paul, 2007: Spatially-variable, physically-derived flash flood guidance. *21st Conf. on Hydrology*, San Antonio, TX, Amer. Meteor. Soc., 6B.2, https://ams.confex.com/ams/87ANNUAL/techprogram/paper_120022.htm.
- Skamarock, W., and Coauthors, 2008: A description of the Advanced Research WRF version 3. NCAR Tech. Note NCAR/TN-475+STR, 113 pp., <http://doi.org/10.5065/D68S4MVH>.
- Smalley, M., T. L'Ecuyer, M. Lebsock, and J. Haynes, 2014: A comparison of precipitation occurrence from the NCEP Stage IV QPE product and the CloudSat Cloud Profiling Radar. *J. Hydrometeorol.*, **15**, 444–458, <https://doi.org/10.1175/JHM-D-13-048.1>.
- Tesfa, T. K., D. G. Tarboton, D. W. Watson, K. A. Schreuders, M. E. Baker, and R. M. Wallace, 2011: Extraction of hydrological proximity measures from DEMs using parallel processing. *Environ. Modell. Software*, **26**, 1696–1709, <https://doi.org/10.1016/j.envsoft.2011.07.018>.
- Van der Vaart, H., 1961: Some extensions of the idea of bias. *Ann. Math. Stat.*, **32**, 436–447, <https://doi.org/10.1214/aoms/1177705051>.
- Wigmosta, M. S., and S. J. Burges, Eds., 2001: *Land Use and Watersheds: Human Influence on Hydrology and Geomorphology in Urban and Forest Areas*. Water Science and Application Series, Vol. 2, Amer. Geophys. Union, 228 pp.
- Wolock, D. M., 1997: STATSGO soil characteristics for the conterminous United States. USGS Open-File Rep. 97-656, <https://doi.org/10.3133/ofr97656>.
- Yang, L., and Coauthors, 2018: A new generation of the United States National Land Cover Database: Requirements, research priorities, design, and implementation strategies.

- ISPRS J. Photogramm. Remote Sens.*, **146**, 108–123, <https://doi.org/10.1016/j.isprsjprs.2018.09.006>.
- Zhang, J., and Coauthors, 2011: National Mosaic and Multi-Sensor QPE (NMQ) system: Description, results, and future plans. *Bull. Amer. Meteor. Soc.*, **92**, 1321–1338, <https://doi.org/10.1175/2011BAMS-D-11-00047.1>.
- , and Coauthors, 2014: Initial operating capabilities of quantitative precipitation estimation in the multi-radar multi-sensor system. *28th Conf. on Hydrology*, Atlanta, GA, Amer. Meteor. Soc., 25 pp., <https://ams.confex.com/ams/94Annual/webprogram/Paper240487.html>.
- Zheng, X., D. G. Tarboton, D. R. Maidment, Y. Y. Liu, and P. Passalacqua, 2018: River channel geometry and rating curve estimation using height above the nearest drainage. *J. Amer. Water Resour. Assoc.*, **54**, 785–806, <https://doi.org/10.1111/1752-1688.12661>.